Seria *Psihologie* este coordonată de Adrian Neculau.

AUREL STAN este lector universitar doctor la Facultatea de Psihologie şi Ştiinţe ale Educaţiei din cadrul Universităţii „Al.I. Cuza", Iaşi. Autor al unor lucrări ştiinţifice din domeniul psihosomaticii, resurselor umane, teoriei testului psihologic şi psihometriei, cu preocupări speciale în computerizarea examenului psihologic şi analiza datelor provenite din investigaţii psihologice. Predă cursuri de statistică psihologică, teoria elaborării testului psihologic şi resurse umane. La Editura Polirom a mai publicat, în colaborare cu Valentin Ciocoţici, *Statistică aplicată în psihologie* (2000).

Cunoştinţele teoretice din domeniul testului psihologic sunt completate de o activitate practică de durată în domeniul psihodiagnosticului, realizată în laboratoare de psihologie din sectoarele industriei, transporturilor şi psiho-pedagogiei speciale. Autorul s-a preocupat de construcţia unor noi teste psihologice computerizate şi de ameliorarea constructivă, prin computerizare, a celor existente în practica psihodiagnostică din România.

www.polirom.ro

© 2002 by Editura POLIROM

Editura POLIROM
Iaşi, B-dul Copou nr. 4, P.O. BOX 266, 6600
Bucureşti, B-dul I.C. Brătianu nr. 6, et. 7, P.O. BOX 1-728, 70700

Descrierea CIP a Bibliotecii Naţionale a României:

STAN, AUREL
Testul psihologic: evoluţie, construcţie, aplicaţii / Aurel Stan – Iaşi: Polirom, 2002.
360 p.; 24 cm (Collegium. Psihologie)
ISBN: 973-683-897-8

159.9.072

Printed in ROMANIA

Aurel Stan

276463

TESTUL PSIHOLOGIC

Evoluţie, construcţie, aplicaţii

035178
B.C.U. - IASI

ACU

POLIROM
2002

159.9.072

606 ex Teste psihologice
x Teoria testului psihologic

29. MAR. 2002

Tablă de materii

Cuvânt înainte 9

Capitolul I

ISTORIA TESTULUI PSIHOLOGIC 11

1. Rădăcinile istorice ale testului psihologic 11

2. De la presistematic la sistematic în construcţia testului psihologic 12

2.1. Contribuţia şcolii psihologice germane la apariţia testului psihologic 15

2.2. Contribuţia şcolii psihologice engleze la apariţia testului psihologic 23

2.3. Contribuţia şcolii psihologice franceze la apariţia testului psihologic 31

2.4. Contribuţia psihiatrilor la apariţia testului psihologic 32

2.5. Comanda socială de psihodiagnostic al caracteristicilor individuale la graniţa dintre secole 36

3. Momentul Binet-Simon în evoluţia testului psihologic 38

3.1. Bazele teoretice, experimentale şi metodologice ale apariţiei scalei Binet-Simon ... 39

3.2. Evaluarea deficienţei individuale 41

3.3. Concepţia lui Binet asupra menirii scalei de inteligenţă 48

3.4. Continuarea principiilor lui Alfred Binet 51

4. Instituţionalizarea psihologiei şi extensia câmpului de aplicare a testelor 51

4.1. Contribuţii americane la dezvoltarea testului psihologic 52

4.1.1. Problema „moronilor" 53

4.1.2. Lansarea coeficientului de inteligenţă înnăscut 58

4.1.3. Reacţia faţă de ereditarism 70

4.1.4. Gigantismul în testarea psihologică 71

4.2. Dezvoltarea mişcării testelor până la al doilea război mondial 81

4.2.1. Dezvoltările conceptuale şi metodologice 82

4.2.2. Concepţia unidimensională şi pluridimensională a inteligenţei 83

4.2.3. Teoria măsurii şi construcţia testelor 85

4.2.4. Diversificarea testelor de inteligenţă şi generalizarea psihometriei 86

4.2.5. Testul nonverbal 92

4.2.6. Testele de aptitudini 92

4.2.7. Tehnicile proiective şi chestionarele de personalitate 93

4.2.8. Inventarele de interese 95

4.2.9. Testele educaţionale 95

4.3. Evoluţia testelor după al doilea război mondial 96
4.3.1. Dezvoltări conceptuale şi metodologice 98
4.3.2. Dezvoltări ale testelor de la începutul anilor '70 până în prezent 100
4.3.3. Evoluţia practicii testelor 101
4.3.4. Intensiv şi extensiv în evoluţia testului 102

Capitolul II
TEORIA TESTULUI PSIHOLOGIC 105
1. Definiţii ale testului psihologic 106
2. Părţile componente ale unui test psihologic 108
3. Clasificarea testelor psihologice 110
4. Teoria clasică a testului psihologic 120
4.1. Precizări 120
4.2. Postulatele teoriei clasice a testului psihologic 121
5. Teoria probabilistă a testului psihologic 126
6. Criteriile de apreciere a unui test 137
6.1. Consideraţii preliminare. Controverse 137
6.2. Criterii principale de apreciere a testului psihologic 139
6.2.1. Obiectivitatea sau concordanţa interpersonală a testului psihologic 139
6.2.2. Fidelitatea testului 140
6.2.3. Validitatea testului psihologic 156
6.2.3.1. Consideraţii generale 156
6.2.3.2. Tipuri de validitate 159
6.2.3.3. Coeficientul de validitate şi interpretarea sa 172
6.2.3.4. Alte tipuri de validitate 193
6.2.3.5. Exigenţe faţă de validitatea unui test 196
6.3. Criterii secundare de apreciere a unui test psihologic 198
6.3.1. Normativitatea unui test 198
6.3.2. Standardizarea unui test 198
6.3.3. Înregistrarea răspunsurilor 200
6.3.4. Evaluarea răspunsurilor la test 200
6.3.5. Interpretarea rezultatelor 200
6.3.6. Comparabilitatea unui test 201
6.3.7. Economia unui test 201
6.3.8. Utilitatea unui test 201
6.4. Relaţiile reciproce între criteriile unui test psihologic 201
7. Analiza de itemi 202
7.1. Itemii ca elemente ale măsurării psihometrice 202
7.2. Tipuri de analize şi indicatori 206
7.2.1. Analiza dificultăţii 206
7.2.2. Indicatori maximali ai itemului 218
7.2.3. Analiza capacităţii de discriminare a unui item 224
7.2.4. Raportul dintre dificultatea şi discriminarea itemului 227
7.2.5. Particularităţi ale indicilor de dificultate şi de discriminare 231
7.2.6. Analiza omogenităţii itemilor 233
7.2.7. Estimarea specificităţii unui test 242

Capitolul III
ASPECTE ALE CONSTRUCŢIEI TESTELOR 245
1. Raţiunea construcţiei unui test 245
2. Etape în construcţia unui test 247
3. Teoria aflată la baza construcţiei testului psihologic 250
4. Formularea unui item bun 253
5. Asigurarea criteriilor de valabilitate a unui instrument psihometric 254
6. Tipuri de solicitări 259
6.1. Tipuri de sarcini cu răspuns legat sau fixat 259
6.2. Sarcini cu răspuns liber 263
6.3. Gradul de complexitate a sarcinii 264
7. Problema eşantionării 265
8. Normarea rezultatelor la testele psihologice 266
8.1. Problema modelelor de măsură 267
8.2. Metode de etalonare a rezultatelor 268
8.3. Tipuri de scale de măsură în psihometrie 273
8.3.1. Scalele de conţinut 273
8.3.2. Scale de dezvoltare 274
8.3.3. Scala pe niveluri de vârstă 274
8.3.4. Scale de nivel şcolar 275
8.3.5. Scale bazate pe raporturi şi coeficienţi 275
8.3.6. Scale bazate pe cuantumuri procentuale sau cuantile 276
8.3.7. Scale de scoruri standardizate 280

Capitolul IV
COMPUTERIZAREA TESTELOR PSIHOLOGICE 285
1. Generalităţi 285
2. Contribuţii aduse computerizării testelor psihologice 286
3. Echivalenţa versiunilor computerizate 294
4. Sisteme de teste computerizate 300
5. Testele adaptative computerizate (CAT) 303
6. Clinic contra actuarial în domeniul psihodiagnosticului 307
7. Preocupări româneşti în domeniul computerizării testelor 308

Capitolul V
CRITICA TESTULUI PSIHOLOGIC 311
1. Critici şi negativişti 311
1.1. Critici din partea psihologilor 312
1.2. Critici din partea altor reprezentanţi ai ştiinţelor sociale 316
1.3. Critici din partea reprezentanţilor societăţii civile 318
1.4. Critici din partea unor organizaţii politice sau autorităţi guvernamentale 319
1.5. Critici din partea beneficiarilor competenţei psihologului 325
2. Dezbateri asupra calităţii testului psihologic în România 327

Capitolul VI
ASPECTE ALE PRACTICII TESTULUI PSIHOLOGIC 331
1. **Domenii de aplicabilitate** 331
1.1. Domeniul psihopedagogic 331
1.2. Domeniul resurselor umane 333
1.3. Domeniul transporturilor 335
1.4. Psihologia clinică 335
1.5. Domeniul psihologiei judiciare 336
1.6. Domeniul cercetării ştiinţifice 336
1.7. Domeniul psihologiei sportive 336
2. **Psihologul şi problemele deontologiei** 337
3. **Teoria şi practica aplicării testelor în România** 340

ÎN LOC DE CONCLUZII 347

Bibliografie 349

B.C.U. „M. EMINESCU" IAŞI

Cuvânt înainte

Testul psihologic a constituit şi constituie încă un domeniu teoretic şi aplicativ diversificat, interesând în primul rând psihologii, dar şi foarte mulţi specialişti din sfera ştiinţelor umane, managementului şi chiar publicul larg. Profunzimea cunoştinţelor privind structura acestuia, dimensiunile analizei sale, diversitatea aplicaţiilor, precum şi implicaţiile folosirii sale în scopuri diagnostice diferă, testul fiind, practic, un instrument specializat de investigaţie. Ignorarea calităţii de domeniu rezervat profesioniştilor, pentru psihodiagnosticul practicat cu ajutorul testului psihologic, a dus la apariţia unor confuzii cu privire la cercul de utilizatori şi modul de exploatare informaţională a rezultatelor sale. Testul psihologic a fost şi a rămas un instrument de investigaţie controversat, cunoscând, de-a lungul evoluţiei sale, oscilaţii în calitatea şi profunzimea aprecierilor, precum şi în interesul acordat de cercetători şi psihologii practicieni.

Pe parcursul acestei cărţi, aducem argumente în favoarea ideii potrivit căreia testul nu este o apariţie „întâmplătoare", ci consecinţa unor acumulări teoretice şi practic-metodologice care şi-au găsit finalizarea la sfârşitul secolului al XIX-lea şi începutul secolului XX, prin activitatea stăruitoare a unor psihologi pasionaţi. Testul este, în primul rând, expresia nevoii de exprimare cantitativă şi calitativă a caracteristicilor psihice implicate în realizarea performanţelor umane, cunoscând ulterior diversificări cerute de necesităţile cercetării şi practicii psihodiagnostice. Această diversificare a făcut ca forma de investigare prin intermediul testului psihologic să fie preluată de un mare număr de domenii ale activităţii ştiinţifice şi practice. A testa potenţialul de cunoştinţe şi abilităţi umane a devenit o practică omniprezentă în civilizaţia contemporană, iar situaţia de examinare prin teste a devenit familiară omului de rând din ţările dezvoltate. Selecţia valorică, însoţită de activităţi de stimulare, perfecţionare şi formare socioprofesională, este un semn distinctiv al lumii civilizate. Ariile geopolitice caracterizate prin dominaţie totalitară au cunoscut contraselecţia valorică, prin împiedicarea circulaţiei oamenilor şi ideilor, prin promovarea unor persoane care nu şi-au confirmat valoarea. Nu întâmplător, în aceste arii, folosirea testelor şi a altor procedee de investigaţie psihologică a fost redusă, uneori chiar interzisă.

Având ca model instrumentul lansat de psihologie, majoritatea disciplinelor ştiinţifice îşi creează teste pentru evaluarea rapidă şi corectă a potenţialului cursanţilor sau candidaţilor la diverse posturi. Considerăm că testul psihologic depăşeşte, prin amploare şi complexitate, orice altă formă de testare.

Nu am neglijat faptul că testarea psihologică a cunoscut contestări venite din partea beneficiarilor competenţei psihologului, a altor specialişti din domeniul ştiinţelor sociale şi a autorităţilor politico-administrative. Am considerat că această contestare trebuie tratată cu atenţie şi luciditate. De multe ori, nu este contestat testul, ci concluzia neconvenabilă, obţinută în urma aplicării unui test. Suntem de părere că aceste contestări reprezintă o sursă de informaţii pentru perfecţionarea şi reevaluarea constructivă a testului.

Prezentarea multiplelor aspecte ale testului psihologic, începând cu etapele evoluţiei, continuând cu teoria structurală şi aspectele tehnice ale construcţiei şi cu prezentarea contestaţiilor la adresa sa, are ca scop sublinierea ideii că testul este o creaţie sistemică, divers condiţionată, care are legi de echilibru specifice domeniilor complex structurate.

Am căutat ca, pe parcursul lucrării, să evităm, pe cât posibil, exprimările tehnicizate şi cantitativiste, dorind să facem, astfel, lucrarea accesibilă unui mare număr de persoane interesate de domeniul abordat. Am putea spune, în acest sens, că nu este o lucrare pură de psihometrie, ci o lucrare asupra problematicii testului psihologic.

Studiul nostru a necesitat o documentare amănunţită şi de durată. Am beneficiat, în acest sens, de generoasele înlesniri ale programului Tempus, în cursul călătoriilor de documentare la biblioteci ale Universităţilor din Paris (X), Liège, Perpignan. Pe lângă acest program, am mai beneficiat de invitaţia primită din partea Auslandsreferat-Universität Konstanz, prin bunăvoinţa doamnei Gerhild Framhein, de donaţiile de cărţi ştiinţifice din partea Institut für Auslandsbeziehungen din Stuttgart şi Inter-Nationes din Bonn. Pe această cale, dorim să mulţumim în mod deosebit instituţiilor amintite pentru generozitatea lor care a făcut posibilă apariţia cărţii de faţă.

Mulţumim, de asemenea, domnului Septimiu Chelcea, profesor la Universitatea din Bucureşti, care a avut bunăvoinţa să lectureze lucrarea înaintea tipăririi şi care a făcut o serie de observaţii juste şi recomandări utile, de care am ţinut cont în completarea formei iniţiale a lucrării. Mulţumirile noastre se adresează şi domnului profesor Valentin Clocotici, de la Universitatea „Al.I. Cuza" din Iaşi, care a avut amabilitatea să verifice corectitudinea formulelor matematice utilizate în prezenta carte.

Considerăm că această lucrare prezintă în primul rând o utilitate practică, verdictul final urmând să fie dat de publicul cititor, care o va aprecia la justa sa valoare.

Sperăm ca lucrarea noastră să fie folositoare atât psihologilor practicieni, cât şi tuturor celor care vor să se iniţieze în problematica testului psihologic. În acelaşi timp, suntem conştienţi de inevitabilele carenţe, deşi considerăm că am realizat câteva punctări care vor contribui la completarea unor aspecte din domeniul cunoaşterii testului psihologic mai puţin abordate în literatura românească de specialitate.

Autorul

Capitolul I

ISTORIA TESTULUI PSIHOLOGIC

1. Rădăcinile istorice ale testului psihologic

Există o tendinţă în rândul istoricilor ştiinţei de a „împinge" începuturile unor discipline ştiinţifice moderne sau relativ moderne până aproape de începuturile civilizaţiei. Nici cercetătorii originii testului psihologic nu scapă de această tentaţie şi susţin că forme rudimentare ale testării pot fi identificate în serviciile publice ale Chinei antice încă din anul 2200 î.e.n. (de exemplu, Teng, în 1942 ; Franke, 1963 ; Lai, 1970 ; Dubois, 1970 ; Chaffe, 1985 ; Benesch, 1995 ; Gregory, 1996). Încă de la data amintită, împăraţii chinezi instituiseră practica examinării funcţionarilor la fiecare trei ani în scopul de a detecta compatibilitatea lor cu posturile pe care le ocupau. Trecerea anilor făcea ca fiecare nouă testare să fie modificată şi finisată. Sursele noastre bibliografice nu explică amănunţit în ce consta această testare imediat după începuturile ei, dar precizează că, începând din 202 î.e.n., dinastia Han a introdus examenul scris la următoarele discipline : legislaţia civilă, problemele militare, agricultura, fiscalitatea şi geografia. Începând din 1370, în confucianismul clasic s-a pus un accent deosebit pe expertiză.

În examinarea preliminară, candidatului i se solicita să petreacă o zi şi o noapte într-o cameră de dimensiuni restrânse şi să compună un poem pe o temă dată. Între 1% şi 7% dintre candidaţi depăşeau această probă şi ajungeau la o instanţă nouă, numită „district de examinare", care instituia 3 sesiuni separate, de 3 zile şi 3 nopţi. Era un examen foarte sever şi foarte riguros, dar nu reprezenta treapta finală. Între 1% şi 10% dintre candidaţii examinaţi se bucurau de privilegiul de a merge la Pekin pentru o fază finală. Exigenţele acesteia erau depăşite de aproximativ 3% dintre participanţi. Câştigătorii din faza finală primeau titlul de mandarin, care însemna „eligibil pentru funcţia de înalt funcţionar public".

Deşi aceste trepte de verificare erau foarte severe, asemănarea cu practica testelor psihologice este superficială pentru cunoscătorii domeniului. Cu toate că procedurile complicate organizate în trepte nu au fost validate, cei care le recomandă în calitate de precursori ai testelor susţin că posedă o serie de criterii relevante. Astfel, scrisul frumos, foarte apreciat aici, era esenţial pentru o comunicare corectă şi concisă. Frumuseţea scrisului şi claritatea exprimării erau considerate predictori relevanţi pentru succesul în funcţiile publice.

În timp, acest complicat sistem de examinare a generat mari nemulţumiri, fiind abandonat în anul 1906, printr-un decret imperial. Trebuie să ne imaginăm cât de mari au fost nemulţumirile, pentru a pune în mişcare ruginitul sistem birocratic chinez.

Peter Hofstätter „sapă" mai adânc în Antichitate şi afirmă că preistoria metodei testului poate fi identificată în riturile de iniţiere ale societăţilor primitive, în care se găsesc aproape totdeauna probe pe baza cărora era apreciată capacitatea unui membru tânăr al tribului de a-şi asuma responsabilităţi specifice adulţilor (*cf.* Hofstätter, 1971, p. 287). Dificultatea principală era dată, în acest caz, de verificarea însuşirilor de caracter (precum curajul şi stăpânirea de sine), a calităţilor raţionamentului (dezlegarea de enigme). Autorul german îşi continuă argumentaţia cu scrierile lui Platon. În operele marelui filosof antic, consacrate statului, sunt descrise anumite procedee de alegere a războinicilor în comunitatea ideală. Calităţile care sunt puse la încercare, de astă dată, sunt autodisciplina, curajul, incoruptibilitatea. Exemple similare celor indicate de profesorul Hofstätter ar putea face subiectul unor tomuri întregi. Mai ales în creaţiile epice ale omenirii, se relatează o mare diversitate de probe. În măsura în care spiritul testului psihologic este spiritul punerii la încercare, se poate spune că acesta traversează istoria civilizaţiei umane.

2. De la presistematic la sistematic în construcţia testului psihologic

Căutarea originilor psihologiei aplicate în timpurile antice poate fi cel mult o exagerare, dar nu o absurditate. Preocuparea pentru studierea psihicului şi a unor metode de diagnostic datează într-adevăr de foarte mult timp, dar ea este inclusă o perioadă apreciabilă de timp în ansamblul strădaniilor filosofilor de a înţelege menirea omului în Univers, de a găsi calea spre adevăr, de a atinge starea de fericire şi soluţia de armonizare a grupurilor sociale cu interese atât de diverse. O serie de tratate de psihologie şi psihiatrie indică originile teoriilor lor în opera lui Platon, Pitagora şi Aristotel, în şcoala stoică, hedonică şi neoplatonică, precum şi în clasificările realizate de medicii Galenus şi Hipocrate. Ce altceva înseamnă „ştiu că nu ştiu nimic" decât chintesenţa teoriei psihologice după care realizarea creativă presupune abandonarea reprezentărilor inadecvate, condiţie esenţială pentru fluidizarea şi flexibilizarea demersului de căutare a noului?[1]

Incursiunile în trecutul preocupărilor psihologice nesistematice ar putea fi analizate pe parcursul mai multor volume. Intenţia noastră este doar de a arăta că psihologia, în calitate de disciplină autonomă, nu a apărut într-un vacuum, ci a fost rezultanta firească a unor acumulări de preocupări nesistematizate sau insuficient sistematizate în plan teoretic şi practic. Apariţia psihologiei ştiinţifice ilustrează o lege a dialecticii lui Hegel, potrivit căreia acumulările cantitative duc la schimbări calitative.

Un ansamblu de cunoştinţe poate purta numele de ştiinţă în momentul în care posedă legi, concepte şi metode proprii, formând un întreg teoretic funcţional. Dobândirea

1. Un inventar sistematic al unor astfel de origini se găseşte în: Petre Brânzei, *Itinerar psihiatric*, Editura Junimea, Iaşi, 1975; Abraham A. Robock, *Welgeschichte der Psychologie und Psychiatrie*, Walter-Verlag, Olten und Freiburg im Breisgau, 1980; Thomas H. Leahey, *A History of Psychology*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1980; William S. Saharian, *History and Systems of Psychology*, FC Peacock Publishers, Inc., Itasca, Illinois, 1968; Daniel H. Robinson, *An Intellectual History of Psychology*, MacMillan Publishing Co., Inc., Collin MacMillan Publishers, Londra, 1970.

statutului de ştiinţă este un drum anevoios, firesc pentru un domeniu de maximă complexitate cum este psihologia.

Cronologic, semnul evoluţiei spre autonomizarea cunoştinţelor psihologice îl reprezintă, fără îndoială, apariţia metodei experimentale. Aceasta este direct dependentă de progresele metodei similare în fiziologie, care, la rândul ei, a fost condiţionată de cea din fizică. Înainte de a exista o psihologie diferenţială, a existat o psihofizică diferenţială şi o psihofiziologie diferenţială. Dintre marii psihologi-fiziologi, îi amintim pe Wundt, Binet, Pavlov, iar dintre marii psihologi-fizicieni pe Fechner şi Helmholtz (*Elemente der Psychophysik* a lui Fechner este, cu siguranţă, prima lucrare de psihologie experimentală). Activarea puternică a unui lanţ de metode a fost posibilă o dată cu descătuşarea epistemologică produsă prin abandonarea concepţiei aristotelice în fizică şi prin ascensiunea unui nou model de gândire, şi anume a celui lansat de Galileo Galilei. Semnificaţia pentru psihologie a acestei treceri a fost analizată în profunzime de către Kurt Lewin, în anul 1931, într-un articol publicat în *Journal for General Psychology*[1].

În lipsa unor profesionişti numiţi psihologi, acumulările de fapte şi concepţii ştiinţifice din secolele XVI-XIX, implicând intervenţia interpretării psihologice, au fost realizate de fizicieni, pedagogi, medici şi, într-o măsură mai mică, de jurişti. Progresul cunoaşterii în faza nesistematică se manifesta prin evidenţierea particularităţilor psihice din cele antropomorfice (fiziognomia), din mişcările expresive, din analiza comportamentului, din observarea procesului de maturizare. Încercările de a perfecţiona aceste procedee de cunoaştere au condus în timp la apariţia testului psihologic. Sistematizarea progresivă a preocupărilor psihologice a fost impusă şi de presiunea socială. Evoluţia instituţiilor sociale a necesitat dezvoltarea unor proceduri psihodiagnostice de diferenţiere a persoanelor care constituiau o subpopulaţie supusă unui interes special. Mărturie stau scrierile medicilor J. Huarte (1575) şi G.B. de la Porta (1673)[2]. Medicul Huarte dă indicaţii moderne asupra diferenţelor de înzestrare, a raportului dintre constituţia corporală şi temperament, a corespondenţei dintre temperament şi înzestrare. Progresul cunoaşterii trebuia să ducă, în concepţia sa, la reformarea învăţământului.

O atenţie specială merită S. Santorio, prieten al lui Galileo Galilei şi fondator al terminologiei medicale, cu studii pe temperamente, în scopul distingerii dintre colerici şi sangvini, consideraţi „calzi" din punct de vedere temperamental, şi flegmatici sau melancolici, consideraţi „reci". El a aplicat pentru prima oară un instrument de măsură în scopuri psihologice (pentru diagnoza temperamentului).

În cunoaşterea fiziognomică şi patognomică se remarcă în cursul secolelor al XVIII-lea şi al XIX-lea J.C. Lavater, J.J. Engel, F.J. Gall, G. Spurzheim, Ch. Bell, C.G. Carus, Th. Piderit, Ch. Darwin şi J.H. Michon. Grafologia contemporană îşi recunoaşte originile în Lavater, Carus şi Michon.

În pedagogie, apar frecvent referiri asupra înzestrării intelectuale ca fundament al organizării şcolare. Revoluţia franceză pune pentru prima dată problema organizării unor

1. Nr. 5, pp. 132-177. De asemenea, vezi „Le conflit dans les modes de pensée aristotélicien et galiléen dans la psychologie contemporaine (1)", in *Psychologie dinamique*, PUF, Paris, 1975, pp. 23-64.
2. Pentru amănunte, vezi Karl Josef Groffmann, „Die Entiwicklung der Inteligenzmessung", in *Handbuch der Psychologie in 12 Bände (6 Band: Psychodiagnostische Diagnostik)*, Verlag für Psychologie, Dr. C. Hogrefe, Göttingen, 1964, pp. 147-190.

clase pentru cei superior dotaţi. În aceeaşi perioadă, pedagogul elveţian Ph.E. von Fellenberg susţine nevoia de consiliere a tinerilor pe baza unui examen psihologic al aptitudinilor.

Adepţii curentului frenologic susţin în scrierile lor existenţa unei accentuate legături între creier şi aptitudinile intelectuale şi afective[1].

În 1798, apare cartea lui Itard, *Băieţii din Aveyron*, unde este amintită pentru prima dată dezbaterea ştiinţifică despre debilitatea mintală, moment important atât pentru istoria psihologiei, cât şi pentru măsurarea inteligenţei. Academia Franceză discută despre experienţa lui Itard de preluare a unor copii găsiţi de vânători într-o pădure şi care trăiseră mai mulţi ani izolaţi de societatea omenească. (O situaţie asemănătoare este descrisă în 1828 de Kaspar Hauser.)

În 1838, J.C. Esquirol (1772-1840) abordează dificila problemă a idioţeniei congenitale şi a demenţei dobândite. El descrie diferite grade de idioţenie, indicând ca mijloc de diagnoză gradul de stăpânire a aptitudinii verbale la idioţi şi, în general, la retardaţi. Metoda asigură şi astăzi o orientare rapidă în cazurile complicate. Procedeul s-a aplicat în multe dintre formele moderne, standardizate, de vocabular ale unor teste de inteligenţă destinate celor normal dotaţi. Esquirol propune distingerea între cei slabi de spirit (cu funcţiile mintale slăbite) şi idioţii propriu-zişi (cu funcţii mintale puţin dezvoltate). Optica sa va fi preluată de Binet.

Esquirol a indicat deosebirea de grad a aptitudinii de randament intelectual şi modul în care se poate recunoaşte aceasta la diferite tipuri de subiecţi.

În 1816, Guggenmoos înfiinţează, în spaţiul cultural german, prima şcoală pentru retardaţi. Necesitatea unor astfel de instituţii este argumentată de psihiatrul Fehring în anul 1821. Unul dintre elevii lui Itard şi Esquirol, şi anume Philippe Seguin, cunoscut pentru formele din scândură utilizate în diagnoza randamentelor intelectuale scăzute, fondează în anul 1837 o şcoală pentru copiii retardaţi din Paris, exemplu urmat de Guggenbuhl (1841) în Elveţia şi de White (1846) în Anglia. „Placa lui Seguin" se menţine, în psihodiagnosticul actual, în cadrul testului Pintner-Paterson. Anul patentării este 1866. Ideile lui Seguin cunosc o rapidă răspândire, după emigrarea lui în anul 1848 în Statele Unite, în statul Massachusetts. (În acest stat, din anul 1846 fiinţa o comisie pentru evidenţierea statistică a retardaţilor.) În anul 1870, existau deja 80 de şcoli private sau de stat pentru retardaţi, copii ori adulţi, în SUA, Canada, Australia şi Europa. Între 1866 şi 1898 numai în Germania existau în jur de 200 de clase ajutătoare cu peste 4.000 de copii. În 1900, von Sickinger organizează în Germania sistemul şcolar Mannheim, ce ia în considerare trei grade de înzestrare, cu clase de copii normal dotaţi, clase cu dotare specială şi clase ajutătoare. Se punea acum accent pe direcţionarea talentelor.

Dezvoltarea practicii intervenţiilor pedagogice a influenţat măsurarea ierarhizării inteligenţei. La sfârşitul secolului al XIX-lea problematica socială şi pedagogică a celor slab dotaţi, precum şi a deficienţilor senzorio-motorii este în general cunoscută. Prin experienţă, s-a putut realiza diferenţierea între idioţenie, dezvoltare normală şi supradotare.

Începutul secolului XX constituie un punct de cotitură, conturându-se foarte clar necesitatea unei mai exacte şi mai rapide deosebiri a capacităţii de randament şi a direcţionării talentelor din şcoală. În Franţa, începând din secolul al XIX-lea, există o puternică dezvoltare a psihopatologiei. Taine Charcot, Bernstein, Liebault, Ribot şi Janet acordă o importanţă deosebită pentru aprecierea normalului şi, firesc, pentru problema individualităţii în psihologie.

1. *Ibidem*, p. 151.

În acelaşi timp, influenţa psihologiei experimentale şi a psihologiei dezvoltării este considerabilă în prefigurarea ideii de *test psihologic*. Ştiinţa, în general, şi experimentarea, în particular, sunt fondate pe observaţii care nu devin precise decât dacă sunt cuantificate. Filosofii apuseni ai secolelor al XVIII-lea şi al XIX-lea aduc în discuţie ideea de măsură în psihologie.

Primul care a vorbit de măsură în psihologie a fost germanul Ch. Wolf care, în lucrările *Psychologia empirica* (1732) şi *Psychologia rationalis* (1734), introduce în limbaj conceptul de *psihometrie*, vorbind de „mărimea plăcerii", „mărimea atenţiei" etc. (*apud* Fraisse, 1967). Ideea de psihometrie este vag evocată de filosofi, naturalişti (Bonod) şi matematicieni (Maupertius, Bernoulli). În 1764, Planquel susţine că gradul de inteligenţă poate fi măsurat, în primul rând, prin numărul de obiecte imaginate, prin intermediul distincţiei cu care obiectele pot fi imaginate, prin intermediul scurtimii timpului necesar pentru a le imagina separat. Un alt contemporan al său (Hagen, 1734) susţinea că atenţia poate fi măsurată prin numărul de idei pe care cineva este capabil să le gândească sau prin timpul în care cineva este capabil de a contempla o idee în toată complexitatea sa, fără discontinuitate. Se descriu diverse experienţe realizabile : a căuta să provoace frică prin descrierea unor pericole iminente, a analiza rezultatele respectivelor stări. Măsurile vizate în descrierea acestor fenomene mintale erau intensitatea şi durata lor. Herbart (1776-1841) considera că psihologia este o ştiinţă care trebuie să se fondeze pe experienţă, metafizică şi matematică, după cum reiese chiar din titlul cărţii sale (*Psychologie als Wissenschaft neu gegrundet auf Erfahrung, Metaphysik und Mathematik*, 1824-1825).

Până la sfârşitul secolului al XIX-lea – consideră Maurice Reuchlin într-o lucrare de istorie a psihologiei (1967) utilizată de noi ca sursă bibliografică – se poate vorbi mai mult de şcoli psihologice naţionale decât de curente psihologice internaţionalizate. Se detaşează clar, în perioada amintită, trei şcoli psihologice bine conturate : germană, engleză şi franceză. Ulterior apare şi şcoala americană, care va marca puternic evoluţia testului psihologic. Psihologiile naţionale menţionate se ocupă în mod inegal de diferenţele individuale.

2.1. Contribuţia şcolii psihologice germane la apariţia testului psihologic

Psihologia germană din perioada corespunzătoare apariţiei testului psihologic era o psihologie preponderent structuralistă şi experimentală. Specificul dezvoltării şcolii germane de psihologie este dat de tradiţia culturală şi ştiinţifică, deosebit de bogată, consistentă şi prestigioasă. Discipline precum filosofia, fiziologia, matematica şi fizica avuseseră şi încă aveau reprezentanţi străluciţi. Intrarea psihologiei în compania onorantă a disciplinelor ştiinţifice şi filosofice din acea perioadă nu se putea realiza mai bine decât prin intermediul reprezentanţilor acestor discipline afirmate. Primii promotori ai psihologiei experimentale germane au fost filosofi, fizicieni, fiziologi, anatomişti etc. care au fost captivaţi de cercetarea psihologică, putând fi consideraţi deschizători de direcţii de cercetare şi creatori ai unui fond peren de idei, în domeniul psihologic. O bună parte din concepţiile lor pot fi considerate actuale şi un număr apreciabil de proceduri experimentale se utilizează şi astăzi. Prima formaţie academică a promotorilor psihologiei experimentale şi ambianţa cultural-ştiinţifică în care au trăit şi-au pus amprenta asupra interpretării rezultatelor obţinute în cercetare. Amintim, ca nişte constante, tendinţa de a fundamenta filosofic concluziile cercetărilor lor, tendinţa de a stabili relaţii cauzale între fizic şi psihic şi tendinţa de a exprima cantitativ rezultatele.

Debutul contribuţiilor la optica experimentală şi structuralistă de abordare a cercetării psihologice pare să fie făcut de Johann F. Herbart, de care am amintit mai sus şi care a aplicat matematica la studiul vieţii psihice, interesându-se de raporturile existente între fenomenele psihice şi fenomenele fizice (pornind de la excitaţia cauzală şi mergând spre stabilirea unor legi).

E.H. Weber (1795-1878), fiziolog şi anatomist, realizează cercetări asupra senzaţiilor tactile şi vizuale, în care îşi propune să urmărească trecerea de la fiziologic la psihologic. Se interesează mai ales de cantitatea de excitaţie necesară pentru ca un subiect să discearnă o primă senzaţie şi, în al doilea rând, de raportul constant şi determinabil care există între senzaţii succesive. Legea lui Weber susţine că excitaţia creşte sau descreşte într-un mod continuu, iar senzaţia discontinuu. Cantitatea de excitaţie corespunzând unui prag diferenţial se află într-un raport bine determinat cu excitaţia de plecare.

Filosoful Gustav T. Fechner (1801-1887) face încercarea unei deducţii matematice mai precise. Preocupările pentru crearea unei ştiinţe psihologice pozitive se aliază, la el, cu o formă curioasă de panteism mistic. Efortul de cercetare al lui Fechner marchează introducerea măsurării în psihologia pragurilor senzoriale. Legea Weber-Fechner a fost mult comentată după apariţia ei. Henri Bergson o critică din punctul de vedere al unui filosof. În esenţă, Bergson susţine că raportul de echivalenţă stabilit între excitaţie şi senzaţia corespunzătoare este o pură convenţie, senzaţiile corespunzând unei impresii subiective, care variază în funcţie de subiect şi chiar la acelaşi subiect în momente diferite, în funcţie de anumite condiţii organice, nervoase, cerebrale. Lucrările lui Weber şi Fechner servesc la determinarea pragurilor senzoriale la anumiţi subiecţi, ceea ce este foarte util, mai ales în orientarea, selecţia şi adaptarea profesională. După aceste prime lucrări ale lui Weber şi Fechner, fenomenele senzoriale au fost obiect de nenumărate studii din perspective foarte diverse (psihofiziologice, gestaltiste, bergsoniene, fenomenologice etc.).

Unii autori consideră tradiţia şcolii experimentaliste germane ca fiind fondată de Gustav T. Fechner şi Hermann von Helmholtz. Paul Fraisse apreciază *Elemente der Psychophysik* a lui Fechner, apărută în 1860, drept prima lucrare de psihologie experimentală (Fraisse, 1967, pp. 1-12). El este preocupat preponderent de raporturile existente între lumea psihică şi cea fizică. Fechner cuantifică într-un mod indirect fenomenele psihice, ceea ce pentru acea vreme putea fi considerat un fapt ştiinţific remarcabil, de excepţie. Henri Bergson consideră această metodă chiar o revoluţie, în ciuda faptului că era un adversar al ideii de măsurare a faptelor psihice (Bergson, 1999). Gustav T. Fechner s-a preocupat să măsoare, cu cea mai mare fineţe posibilă, pragurile diferenţiale, dezvoltând metode de măsură care sunt încă utilizate în cercetarea percepţiei. Dintre metodele de măsurare care pot fi recunoscute ca utilizate şi astăzi, amintim:

- metoda erorii medii, cunoscută astăzi sub numele de *metoda de ajustare sau de reproducere*;
- metoda cazurilor adevărate şi false, numită şi *metoda constanţei*;
- metoda diferenţelor just perceptibile, cunoscută sub numele de *metoda limitelor*.

Actualmente, antrenamentul cu aceste metode constituie exerciţii pentru formaţia de bază a psihologilor experimentalişti. Calităţile care le menţin în actualitate sunt următoarele:

- sunt esenţial psihologice şi iau în considerare dificultăţile proprii oricărei determinări cantitative;

- recunosc importanţa variabilităţii estimărilor momentane şi utilitatea recurgerii la medii;
- ţin cont de distorsiunile pe care le antrenează ordinea estimărilor şi de necesitatea de contrabalansare.

Gustav Fechner aplică la erorile de măsură în psihologie *legea Gauss-Laplace*, model de distribuţie cu aplicabilitate chiar în domenii de cercetare de o amplă diversitate. Cu această *lege*, pătrunde în noua ştiinţă a psihologiei ideea de probabilitate, cu aplicaţii foarte largi.

Un alt savant care a contribuit la progresul teoriei experimentale şi care a avut idei sclipitoare în acest domeniu este Hermann von Helmholtz, menţionat mai sus. Formaţia academică principală ţine de domeniul fiziologiei, unde are realizări de excepţie, dar aportul său în psihologie este la fel de important. A inventat aparate de experimentare şi metode de măsură, dintre care amintim metoda ce utilizează timpul de reacţie pentru a măsura influxul nervos la om. De asemenea, elaborează o teorie a percepţiei numită a *inferenţelor inconştiente*, pe care o opune teoriilor nativiste, susţinând ideea potrivit căreia în percepţiile noastre există mult mai mult decât datele senzoriale actuale. Acest *mai mult* vine din inferenţele inconştiente irezistibile care se formează prin experienţă şi în care asociaţia şi repetiţia joacă un rol esenţial. Helmholtz trage o concluzie foarte importantă în privinţa valorii observaţiilor ştiinţifice care depind de experienţa trecută a experimentatorului. Pentru savantul german, în activitatea de cercetare, este prezentă întotdeauna o ecuaţie personală a savantului, fiind necesară o *atmosferă de laborator* ce antrenează cercetătorul în observarea mai curând a anumitor fenomene decât a altora.

Primele laboratoare de psihologie experimentală sunt înfiinţate în Germania de către Wilhelm Wundt (Leipzig, 1879) şi Georg Elias Müller (Göttingen, 1881). Numărul de laboratoare înfiinţate poate fi considerat un indicator al amplorii cercetărilor experimentale. În perioada 1879-1896, s-au înfiinţat în lume un număr de 57 de laboratoare de psihologie experimentală, şi anume 31 în SUA, 10 în Germania, 3 în Rusia, 3 în Italia, 3 în Franţa şi câte unul în Anglia, Danemarca, Japonia, Belgia, Olanda, Austria, Canada. 47 dintre aceste laboratoare au fost înfiinţate în cadrul unor universităţi.

Personalitatea care se detaşează net, în privinţa contribuţiilor la apariţia psihologiei ştiinţifice, este Wilhelm Wundt (1832-1920), care lansează, în domeniul cercetării ştiinţifice de anvergură, psihologia experimentală ca psihologie fiziologică, deoarece problematicile şi metodele sunt apropiate fiziologiei şi, în general, ştiinţelor naturale. Wundt îşi propune să elaboreze o psihologie în care să nu admită decât „fapte", recurgând în măsura posibilului la experiment şi la măsură.

În concepţia lui Wundt, psihologia, pentru a putea exista ca ştiinţă naturală, trebuie să răspundă aproape în întregime la câmpul cercetărilor experimentale şi la cel al măsurii. Experienţa este mijlocul esenţial care conduce faptele de conştiinţă spre acel fundal obscur în care se elaborează viaţa conştientă. Observaţia interioară, ca orice observaţie în general, nu ne oferă decât fenomene compuse. Prin experimentare, despuiem fenomenul psihic de toate condiţiile accesorii. Prin experimentare, producem fenomenele artificial, în condiţii determinate, pe care putem să le schimbăm după voinţa noastră şi care ne permit să le măsurăm. Pretutindeni, experienţa conduce la legi naturale, pentru că ea ne permite să sesizăm simultan cauza şi efectul. Experienţa este însoţită pas cu pas de către măsură. A măsura şi a cântări – acestea sunt marile mijloace de care cercetările experimentale se servesc aproape întotdeauna pentru a ajunge la nişte legi precise. Toate

măsurările pot să traducă rezultatele lor prin numere, care nu sunt scopul măsurii, ci mijloace indispensabile pentru a ajunge la scopul ultim al cercetării.

Interesul ştiinţific se manifestă în privinţa proceselor senzoriale şi proceselor perceptive elementare. Totuşi, şcoala experimentalistă creată de Wundt nu respinge speculaţia filosofică şi practică introspecţia, metodă respinsă de mulţi experimentalişti ca subiectivă. Când ea pretinde să se intereseze de procesele superioare, acestea sunt asimilate, în fapt, celor mai elementare procese (de exemplu, inteligenţa va fi asimilată rapidităţii timpului de reacţie). Aceste procese elementare sunt considerate elemente simple care se cer definite cu precizie şi în cazul cărora se încearcă punerea în evidenţă a conexiunilor în cel mai pur spirit asociaţionist.

În laboratoarele lui Wundt sunt măsuraţi indivizii, dar nimeni nu se preocupă de a descrie variabilitatea interindividuală, nici de a o explica, aceasta fiind considerată o sursă de eroare aleatorie. Prin intermediul măsurilor efectuate, se încearcă stabilirea unor legi generale, valabile pentru toţi indivizii. Această psihologie experimentală a contribuit la apariţia testelor. James McKeen Cattell (1860-1944) a relevat importanţa fenomenelor de variabilitate interindividuală pe când studia, în coordonarea lui Wundt, factorii care contribuie la modificarea timpului de reacţie. Standardizarea observaţiei ce caracterizează testele reprezintă un aport al metodologiei experimentale germane. Primele teste nu vor fi, în multe cazuri, decât adaptări ale dispozitivelor experimentale din laborator. Wilhelm Wundt, James McKeen Cattell şi Francis Galton pun fundamentul pentru testele psihologice ale secolului XX (după cum se exprimă plastic Robert Gregory, 1996, p. 5).

În anul 1862, într-o lucrare asupra percepţiei senzoriale, W. Wundt propune constituirea oficială a unei noi ştiinţe, psihologia experimentală. În 1874, publică *Grundzüge der physiologischen Psychologie* (*Principii de psihologie fiziologică*), considerat primul tratat complet de psihologie.

Wundt abandonează preocupările fiziologice de la Zürich şi pleacă la Leipzig, pentru a fonda, în 1879, primul Institut de Psihologie din lume care posedă un laborator de psihologie experimentală. Lucrările primului laborator de psihologie din Leipzig sunt publicate în revista *Philosophisches Studien*, creată de Wundt, în anul 1881. Tânăra disciplină nu era considerată o nouă filosofie, nici o negaţie a filosofiei, ci o dezvoltare, un progres al gândirii filosofice.

Constituirea testului psihologic modern datorează la fel de mult psihiatrilor şi celor care au înfiinţat laboratoare de psihologie experimentală. Un merit incontestabil al lui Wilhelm Wundt constă în aceea că a atras un important număr de tineri studioşi, dotaţi, care au răspândit în toată lumea ideile şi tehnicile experimentale. Îi amintim pe E. Kraepelin, O. Külpe, E. Meumann (Germania), G.S. Hall, J. McKeen Cattell, E.B. Titchner (SUA), Ch. Spearman (Anglia), Bourdon (Franţa), Thierry şi Michotte (Belgia). Sistematizarea demersului experimental este o puternică trăsătură a şcolii create de Wundt. Cercetările aveau un conţinut clar şi s-a acordat o deosebită importanţă dominării condiţiilor care intervin în experiment. De asemenea, se bucură de o atenţie aparte condiţiile exterioare, prezentarea stimulilor, ordinea încercărilor, care pot influenţa într-un grad înalt rezultatele. Problema standardizării este prezentă în conştiinţa celor care cercetau fenomenele psihologice. S-au căutat mai ales proceduri obiective, capabile de replicare. Aceste exigenţe au reprezentat condiţii metodologice premergătoare apariţiei testului psihologic, care constituie instrumentul de bază al psihologiei diferenţiale.

Psihologia experimentală dezvoltată în Germania şi Anglia reuşeşte să depăşească metodele subiective ale introspecţionismului. Totuşi, accentuarea metodelor obiective, precum şi natura calităţilor măsurate duc, în unele cazuri, la improvizare, accentuată şi de o concepţie sterilă. Psihologia experimentală timpurie a greşit considerând că procesele senzoriale simple pot fi folosite pentru măsurarea inteligenţei.

Pentru Wundt, psihologia experimentală trebuie să fie o ştiinţă a *faptelor interne* şi o fizică a *faptelor externe*. În concepţia lui, problema generală a psihologiei experimentale se referă la trei interogaţii:

- Care sunt elementele conştiinţei, proprietăţile sale calitative şi cantitative?
- Cum se combină ele pentru a forma fapte de conştiinţă, de o natură din ce în ce mai complexă?
- Care sunt raporturile de coexistenţă şi de succesiune, legile generale care domină faptele interioare?

În primul rând, Wundt realizează studii asupra percepţiei vizuale, respingând ipoteza energiei senzoriale specifice. În cercetări, îşi propune să determine importanţa fenomenelor psihice cu substratul lor organic, în particular cerebral. Presupune existenţa a două tipuri de legi ale cunoşterii: legi asociative şi legi aperceptive (care exprimă activitatea liberă a gândirii). Influenţat de Immanuel Kant, el refuză orice validitate noţiunilor decurgând dintr-o concepţie substanţialistă a sufletului, admiţând că aperceptiile joacă un rol fundamental în actul cognitiv. Apercepţia activă, care are drept corolar voinţa, este singura garanţie a continuităţii interioare. Un rol fundamental, în viaţa psihică, este acordat instinctului.

W. Wundt consideră că între organizarea fizică şi operaţiile psihice există o relaţie de cauzalitate, în care prima enumerată este privilegiată. Modul în care înţelege relaţiile între diversele elemente ale structurii psiho-fizice a fiinţei umane ajunge la o formă de metafizică, în care concepţia animistă este considerată cea mai bună bază pentru problema dezvoltării fizice şi intelectuale. Admite, de asemenea, o anumită finalitate la toate fenomenele naturii (*apud* Nicolas, 1999, p. 27).

În privinţa atitudinii faţă de contribuţiile lui Fechner, Wundt consideră că cele două proprietăţi obiective ale senzaţiei sunt intensitatea şi calitatea. Un lucru este cunoaşterea senzaţiilor, un altul este urmărirea combinaţiilor lor pentru a forma faptele de conştiinţă. Psihologia experimentală urmează un demers lent şi progresiv, începând prin studierea combinaţiilor celor mai simple, şi anume percepţiile. Wundt admite că, într-un anumit caz, percepţia este analoagă unei combinaţii chimice. Senzaţiile elementare pot fi izolate, dar nu se regăsesc în elemente proprietăţile compusului pe care îl formează prin unire. Astfel, o percepţie este formată din două sau mai multe senzaţii pure, combinate prin mecanismele de asociere care permit construirea unor sisteme de reprezentări ale lumii exterioare. Mecanismele de asociere sunt, în general, de natură automată, dar subiectul poate să etaleze mecanismele de asociere activă prin intervenţia proceselor *aperceptive*, declanşate de atenţie (în opera lui Wundt, conceptele de *atenţie* şi de *apercepţie* sunt strâns legate). Atenţia este o stare mintală caracterizată printr-un sentiment special care însoţeşte sesizarea netă a unui conţinut, pe când apercepţia este un proces voluntar activ care permite această sesizare. În scopul de a cerceta elementele constitutive ale unei percepţii, ca şi legile lor de combinare, pot fi utilizate mai multe metode. Aceea pe care Wundt o va privilegia constă în a opera modificări în calitatea unei percepţii sau a condiţiilor sub care ea se prezintă, în scopul evidenţierii elementelor sale constitutive. Acestea sunt

tehnici de bază ale psihologiei experimentale, care constau în varierea modalităţilor unui factor experimental, pentru a-i observa efectele asupra comportamentului.

Pentru a măsura durata actelor psihice, Wundt utilizează două clase de metode experimentale:

- *metodele de reacţie*;
- *metodele de comparare*.

Metodele de reacţie constau în determinarea duratei unui fapt relativ simplu, dar care, în realitate, este foarte complex, durată numită de către Wundt reacţie simplă. La o anumită excitaţie, subiectul trebuie să răspundă în cel mai scurt timp posibil, printr-un mod de acţiune prestabilit. Timpul cerut de acest act de reacţie este determinat cu ajutorul unui instrument cronometric (cronoscopul lui Hipp) care înregistrează momentul de excitaţie şi momentul răspunsului dat de subiect.

Prin intermediul *metodelor de comparare* se adaugă, la ansamblul faptelor fizice şi psihice care cuprind reacţia simplă, anumite fapte psihice cunoscute. Apoi, se măsoară durata unui astfel de fapt psihic compus şi se scade din durata sa totală durata reacţiei simple, obţinându-se în acest fel durata actului adiţional. Deoarece fenomenele psihice complexe cuprind fenomenele cele mai simple, se poate complica progresiv actul psihic şi, astfel, se poate obţine durata fiecărui act adiţional, prin scăderea succesivă a duratei necesare a actelor mai simple care formează actul total. Este un procedeu care poate fi apreciat ca o primă încercare de măsurare a faptelor psihice, dar care considerăm că este deficitar prin simplismul său psiho-fizic.

Datorită conţinutului metalului folosit în construcţia aparatelor de măsură, etapa temporală respectivă din istoria psihologiei experimentale este numită *era instrumentelor de alamă a examinării psihologice.* Aceste instrumente erau adaptate pentru a măsura pragurile senzoriale şi timpul de reacţie (psihologii vremii considerau că, prin cercetarea acestor abilităţi, se află în centrul inteligenţei). În ciuda startului fals dat de experimentaliştii timpurii în cercetarea proceselor psihice superioare, aceştia au meritul de a fi furnizat psihologiei o metodologie mai adecvată ale cărei componente persistă, parţial, în cercetarea psihologică actuală. Faptul că între rezultatele obţinute nu era o unitate deplină a adus în prim-plan cercetarea diferenţelor individuale, direcţie urmată de către James McKeen Cattell. În 1882, psihiatrul Emil Kraepelin evidenţiază în lucrările sale ştiinţifice importanţa diferenţelor individuale.

Hermann Ebbinghaus (1850-1909) poate fi revendicat şi de alte şcoli naţionale de psihologie, având în vedere faptul că el a activat şi în Franţa şi Anglia. Este unul dintre primii psihologi care au abordat funcţiile superioare ale psihicului, folosind metode experimentale. În cursul formării sale ştiinţifice, este foarte impresionat de lucrarea *Elemente der Psychophysik*, de unde extrage sugestia studierii unor domenii mai complexe ale psihicului, precum memoria. Contrar lui Herbart şi psihologilor experimentalişti din vremea sa (de exemplu Wundt), el gândeşte că o ştiinţă experimentală a proceselor mintale superioare ar fi posibilă. A fost, de asemenea, unul dintre primii care au luat în considerare expresia inconştientă a fenomenelor psihice. Ebbinghaus nu reduce memoria la amintirea conştientă, ci decide dezvoltarea unui indicator bazat pe economia în timp sau în randamentul unei noi încercări.

Metoda utilizată, denumită *economia la reînvăţare*, avea avantajul de a permite abordarea acestei funcţii psihologice, fără a reduce la fel de mult expresia sa conştientă. Pentru a favoriza studiul obiectiv al memoriei, Ebbinghaus decide să introducă un

material nou (serii de silabe fără semnificaţie, din care realizează aproximativ 2.300), procedeu care reduce pe cât posibil influenţa semnificaţiei, precum şi metode cantitative susceptibile de a sprijini concluziile sale. Fără ajutor şi fără a avea un laborator propriu, el dezvoltă o lungă serie de investigaţii experimentale în domeniul memoriei, ale căror rezultate au fost publicate într-o operă de sinteză, *Über das Gedächtnis : Untersuchungen zur experimentelle Psychologie* (*Asupra memoriei : cercetări de psihologie experimentală*), în anul 1885. Opera lui Ebbinghaus va declanşa o serie de cercetări asupra memoriei şi învăţării în Germania, Franţa şi Statele Unite. Graţie acestei lucrări, este promovat profesor la Universitatea din Berlin, unde fondează, în 1886, un laborator de psihologie. Studiile sale sunt publicate, în majoritate, într-o revistă pe care o va înfiinţa în 1890 împreună cu Arthur König, *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* (*Revista de psihologie şi fiziologie a organelor de simţ*), devenită ulterior *Zeitschrift der Psychologie*, publicaţie deosebit de prestigioasă care apare şi astăzi în Germania. Revista patronată de Ebbinghaus accepta lucrări de istoria psihologiei, de psihologie experimentală şi metodologie.

În realizarea cercetărilor, a avut ideea de a utiliza numărul de repetiţii ca măsură a memoriei. Inventează *silabele fără sens* ca material de studiu. Cu acest material şi cu poezii ajunge la metodele, devenite clasice, de studiu al memoriei, metoda de învăţare completată cu metoda de economie. Se preocupă prioritar de:

- influenţa lungimii materialului de învăţat;
- sarcina de memorizare în funcţie de numărul de repetiţii;
- uitarea şi curba sa în funcţie de timp.

În studiile sale, Ebbinghaus proclamă necesitatea absolută a introspecţiei în activitatea de cercetare. *Über das Gedächtnis*, model de claritate şi precizie, probează faptul că psihologia poate să iasă din domeniul cercetării senzaţiilor cu condiţia de a inova. În 1897, elaborează *metoda completării*, pentru a studia capacitatea mintală a şcolarilor, studiu care a fost cerut de magistraţii municipali din Breslau, în vederea unei mai bune distribuţii a orelor de activitate a copiilor. Probele create pentru aceste studii sunt utilizate şi astăzi drept componente ale unor teste de inteligenţă (a se vedea WISC).

Şcoala de la Würzburg. Această şcoală a apărut ca urmare a dorinţei unora dintre elevii lui Wundt de a ieşi din studiul exclusiv al senzaţiilor, în care s-a cantonat maestrul lor, şi de a aborda direct cercetările experimentale ale funcţiilor superioare (teren interzis experimentării, după principiile lui Wundt). Principalul promotor al acestei tendinţe de cercetare a fost asistentul lui Wilhelm Wundt, Oswald Külpe. Primele tentative de distanţare a acestuia faţă de maestrul său sunt făcute în lucrarea publicată sub titlul *Grundriss der Psychologie* (*Compendiu de psihologie*), după ce timp de doisprezece ani a continuat activitatea în spiritul lui Wundt. Külpe a fost unul dintre puţinii reprezentanţi ai şcolii germane de psihologie de la sfârşitul secolului al XIX-lea, format în propria disciplină (a studiat istoria şi psihologia). În cartea amintită anterior, publicată în 1893, a făcut jurământul de credinţă de a nu vorbi decât despre *fapte ştiinţifice*, oricare ar fi lacunele pe care această orientare le presupune. Din anul 1894, este profesor la Würzburg, unde creează propriul laborator, şi apoi, în 1896, Institutul de Psihologie, împreună cu un vechi elev al lui Wundt, Karl Marbe. Ceea ce distinge Şcoala de la Würzburg este teza fundamentală potrivit căreia este posibil, prin intermediul introspecţiei, să se studieze experimental funcţiile psihice superioare. Partizanii introspecţiei ajung la concluzia că în

lumea psihică accesibilă autoobservaţiei nu există doar imagini, ci şi operaţii, unii dintre ei afirmând, nu fără a fi combătuţi de către alţi psihologi, că există şi operaţii mintale neînsoţite de imagini (gândire fără imagini). Şcoala introspecţionistă se pune astfel în opoziţie deschisă faţă de asociaţionism. Departe de a renunţa la metodele psiho-fizice, Şcoala de la Würzburg le foloseşte pentru a provoca autoobservarea. Külpe şi colaboratorii întreprind, în prima jumătate a secolului XX, explorarea experimentală a regiunilor superioare ale lumii psihice (abstractizarea, judecata, gândirea, raţionamentul). Această şcoală afirmă existenţa unui conţinut al gândirii ireductibil la senzaţie şi, în general, a unei activităţi psihice ireductibile la senzaţii, care este mai complexă decât mecanismele de asociaţie. Ulterior, Oswald Külpe se manifestă ca un adversar al lui Wundt şi consideră că se poate experimenta asupra proceselor superioare, încercând să facă pentru gândire ceea ce Ebbinghaus a reuşit asupra memoriei. Eşuează în încercarea sa, dar nereuşita va genera o schimbare de optică în cercetarea psihologică. Psihologia este ştiinţa faptelor, care depind de experienţa subiectului. În experienţele sale, apelează frecvent la studenţi, dar rezultatele obţinute sunt puţin relevante. Pentru a pune bazele unei introspecţii experimentale sistematice, propune o mai bună analiză, prin fracţionarea momentelor unei operaţii mintale şi prin susţinerea ideii că totul trebuie să fie fundamentat pe experienţă.

Gestaltismul, ca teorie a preeminenţei totalităţii asupra părţilor, reprezintă în domeniul percepţiei un alt curent de gândire iniţiat în spaţiul cultural german, care cuprinde şi o reacţie contra elementarismului psihologilor asociaţionişti şi care poate fi considerat ca ţinând de psihologia fenomenală. Ideile fundamentale ale psihologiei formei fac parte din orice cultură psihologică.

Acest curent era abia conturat în perioada anterioară creării testului psihologic, el dezvoltându-se, ulterior, cu influenţă benefică asupra bazei teoretice a tehnicilor proiective. Originea sa o găsim în aşa-numita şcoală austriacă (*Aktpsychologie*), compusă din elevii lui Fr. von Brentano, dintre care îi enumerăm pe A. Meinong, Chr. von Ehrenfels, St. Witasek, V. Benussi etc.

Christian von Ehrenfels (1859-1932) este primul care distinge în percepţie calităţile formale (*Gestaltqualitäten*) şi care susţine cu tărie ideea potrivit căreia întregul reprezintă mai mult decât suma părţilor componente. În consecinţă, percepţia este mai mult decât o fuziune a două senzaţii elementare. Un precursor al acestui nou mod de gândire poate fi considerat R.H. Lotze (1817-1881), care elaborează, în 1852, *teoria semnelor locale*. Fiecare senzaţie, în afara calităţilor particulare, va avea caractere intensive proprii, legate de localizare. Spaţiul naşte legături între aceste semnale locale prin experienţă şi mişcare. Încă din 1886, E. Mach consideră că spaţiul şi timpul contribuie la apariţia unor senzaţii originale şi crede că există cu adevărat un simţ al timpului.

Christian von Ehrenfels susţine că spaţiul şi timpul sunt calităţi noi, care nu rezultă dintr-o simplă combinaţie de elemente. Forma poate să rămână, când elementele îşi schimbă natura. El susţine, de asemenea, că proprietăţile formale nu rezultă din relaţia dintre elemente, ci din activitatea spiritului care combină respectivele elemente. Noţiunea de *calitate formală* apare în perspectiva deschisă prin psihologia actului.

Max Wertheimer (1880-1943), ce poate fi inclus şcolii austriece, consideră mişcarea stroboscopică ca fiind o experienţă exemplară pentru relaţia întreg – parte. Mişcarea, pe care o denumeşte *fenomenul phi*, depinde de structura de ansamblu şi, în particular, de numărul de elemente. Întregul nu este suma părţilor, iar anumite aspecte rezultate nu pot fi deduse din părţi.

Ideile ştiinţifice ale şcolii germane vor fi preluate de psihologii care au creat probe psihologice, mai ales de către cei care au introdus rigoarea în activitatea lor şi care au căutat să aibă o bună fundamentare teoretică a noţiunilor folosite. Tehnicile proiective îşi găsesc fundamentul teoretic în lucrările reprezentanţilor psihologiei formei.

2.2. Contribuţia şcolii psihologice engleze la apariţia testului psihologic

Psihologia engleză de la sfârşitul secolului al XIX-lea şi începutul secolului XX a fost, prin excelenţă, o psihologie diferenţială şi va contribui cel mai mult la apariţia testelor.

Fondatorul psihologiei moderne engleze este, indubitabil, Francis Galton (1822-1911), datorită capacităţii sale de a pune în aplicare cunoştinţele şi metodele psihologice într-un mod sistematic şi de a constitui ansambluri vaste de date în domeniul cercetării psihologice. În Anglia, serviciile sale aduse progresului cunoştinţelor psihologice pot fi asemănate cu cele ale lui Wilhelm Wundt în Germania. E.G. Boring consideră că Wundt este fondator al psihologiei generale, a adulţilor, a normalilor, iar Galton – al psihologiei diferenţiale, individuale, a psihicului normal.

Este influenţat în formarea sa intelectuală de curentul evoluţionist. B.R. Hergenhahn, istoric al psihologiei, consideră curentul psihologic care promovează testele ca fiind influenţat în mare măsură de teoria darwinistă. Capitolul destinat pionierilor testului psihologic este intitulat, în lucrarea sa consacrată istoriei psihologiei, „The Darwinian Influence" (Hergenhahn, 1997, pp. 263-300).

Sir Francis Galton este una dintre personalităţile exemplare ale istoriei ştiinţei universale. S-a născut la Birmingham într-o familie aristocratică bogată. Erasmus Darwin, fizician, filosof, poet şi teoretician al evoluţionismului timpuriu, a fost bunicul lui Charles Darwin şi Francis Galton. Tatăl său era un bancher bogat şi mama sa era soră cu tatăl lui Charles Darwin. Educaţia timpurie o primeşte acasă. Numele lui Galton este folosit de multe ori ca exemplu de autorii care se ocupă de copiii precoce şi de cei care studiază culmile potenţialului intelectual uman. Începe să scrie şi să citească la 2 ani şi jumătate, la 5 ani poate să citească în engleză câteva cărţi, iar la 7 ani citeşte din pasiune autori ca Shakespeare. Este trimis la o şcoală particulară cu o atmosferă educativă bazată pe îngrădirea libertăţii elevului, privaţiune şi pedepse fizice. La 16 ani se eliberează de povara acestei instituţii de învăţământ şi este trimis la Birmingham General Hospital pentru a studia medicina. După această experienţă practică, este transferat la King's College din Londra. După mutarea la Cambridge University, unde obţine diploma universitară în 1843, Galton îşi planifică case să se reîntoarcă la King's College pentru a obţine o diplomă în medicină. Moartea tatălui său îl determină să renunţe la acest proiect şi să-şi încheie educaţia formală. Fiind un aristocrat bogat, Galton are posibilitatea să facă nenumărate studii şi cercetări, total degajat de grijile materiale, lucrând unde dorea şi când dorea. După obţinerea titlului universitar, a călătorit în Egipt, Sudan şi Orientul Mijlociu. Întors acasă, petrece un timp în societatea bogaţilor săi prieteni, organizând partide de vânătoare, călărind, călătorind cu balonul şi efectuând experimente în domeniul electricităţii. După consultarea unui frenolog, care-i recomandă o viaţă activă, Galton

decide să intre în Royal Geographical Society şi face o călătorie ştiinţifică de doi ani în sud-vestul Africii. Realizează cu acest prilej o hartă cu teritoriile nou explorate din Africa, pentru care Royal Geographical Society îl onorează, în 1853, la vârsta de 32 de ani, cu cea mai înaltă medalie. În abilitatea cu care Galton face hărţi, se poate vedea pasiunea lui de o viaţă de a măsura lucrurile.

În 1853, Galton publică prima sa carte, *Narrative of an Explorer in Tropical South Africa*, pentru care i se recunoaşte calitatea de expert în călătoriile în zone sălbatice. Totodată, guvernul britanic îl împuterniceşte să înveţe soldaţii în proceduri de campare. În 1855, publică a doua carte, *The Art of Travel*, care aduce informaţii despre fauna şi locuitorii zonelor mai puţin cunoscute ale continentului african. Pentru contribuţiile sale în domeniul explorărilor geografice Francis Galton a fost ales preşedinte al Royal Geographical Society în anul 1856. Îşi manifestă creativitatea şi în domeniul meteorologiei şi criminalisticii. În efortul de a măsura şi prezice vremea, inventează harta meteorologică, fiind primul om de ştiinţă în acest domeniu care foloseşte termenii *minimă*, *maximă şi front*. De asemenea, a fost primul care a sugerat că amprentele pot fi folosite pentru identificarea persoanelor, procedură preluată mai târziu de Scotland Yard.

Dintre lucrările sale ulterioare, în bună parte consacrate problemelor de psihologie diferenţială, amintim:

1869 – *Hereditary Genius : An Inquiry into its Laws and Consequences* ;
1874 – *English Men of Sciences : Their Nature and Nurture* ;
1875 – *The History of Twins as a Criterion of the Relative Powers of Nature and Nurture* ;
1883 – *Inquiries into Human Faculty and its Development* ;
1888 – *Co-relations and Their Measurement, Chiefly from Anthropometric Data* ;
1889 – *Natural Inheritance*.

Pasiunea lui Galton pentru măsurători cunoaşte, ulterior, şi alte realizări.

Teoria evoluţionistă este reprezentată în psihologie de opera lui Charles Darwin, care, în 1877, introduce metoda biografică a copilului, preluată apoi ca metodă fundamentală a psihologiei dezvoltării, după ce W. Preyer face o prezentare sistematică într-o carte care cunoaşte până în 1900 cinci ediţii internaţionale[1]. Prin americanul G.St. Hall, în anul 1883, îşi găseşte utilizare şi răspândire rapidă metoda chestionarului la copii, care realizează totodată şi consideraţii statistice.

Francis Galton îşi propune să aplice teoria evoluţionistă la specia umană, primeşte cu bucurie teoria diferenţelor individuale a vărului său şi consideră că misiunea vieţii sale este să măsoare aceste diferenţe.

O asemenea aplicare acordă un rol central diferenţelor individuale, deoarece ele se află în centrul teoriei darwiniste. Darwin considera că există o variabilitate interindividuală, care se explică parţial prin ereditate. Proiectele lui Galton cuprind trei componente :

- în primul rând, îşi propune să descrie variabilitatea interindividuală. O dată cu această preocupare, pune bazele statisticii descriptive şi metodologiei testelor (introduce noţiunile de *etalonare*, *regresie*, *corelaţie*) ;

1. Lucrarea a apărut în 1881 la Leipzig şi se intitula *Die Seele des Kindes* (*Sufletul copilului*).

- în al doilea rând, arată că diferenţele individuale sunt influenţate de ereditate. Pentru a crea un suport metodologic teoriei diferenţelor individuale, dezvoltă metoda arborelui genealogic şi inventează aşa-numita *metodă a gemenilor*. În scopul de a demonstra caracterul ereditar al diferenţelor individuale, defineşte coeficientul de corelaţie, care va fi perfecţionat de alţi reprezentanţi ai şcolii psihometrice engleze. Doreşte să descopere legile eredităţii, măsurând şi analizând gradul de asemănare între părinţi şi copii ;
- în al treilea rând, îşi propune să lupte contra degradării speciei umane. Galton consideră că specia umană este în pericol şi se angajează în lupta eugenistă (este primul care utilizează acest termen, în anul 1863).

Teoria cuprinsă în *Hereditary Genius* susţine ideea conform căreia cei superior dotaţi, posedând calităţi speciale, provin dintr-o anumită linie ereditară şi aparţin unei anumite familii. G. Murphi considera că avem de-a face cu o primă tratare sistematică a problemei diferenţelor individuale. Lucrarea lui Galton, *Inquiries into Human Faculty and its Development*, dedicată diferenţelor individuale ale facultăţilor mintale, este considerată de Boring, în 1950, începutul mişcării testelor mintale, anunţând apariţia psihologiei ştiinţifice a diferenţelor individuale. Cartea, o curioasă mixtură a cercetării empirice cu încercări speculative pe teme diverse, este consacrată diferenţelor clar perceptibile. El consideră important nu numai faptul că diferenţele individuale există, ci şi că ele pot fi măsurate. De asemenea, adaptează procedurile existente în Europa, perfecţionând măsurările senzorio-motorii. Continuă tradiţia testelor mintale a *instrumentelor de alamă*, cu o importantă diferenţă : colecţia de date cuprinde serii de sute şi mii de subiecţi. Galton face consideraţii şi comparaţii statistice, în lucrările apărute în 1869, respectiv 1883, iar examenele sale prin teste a avut o mare influenţă asupra evoluţiei acestei proceduri în SUA. Semnificaţia cuvântului „test", ca probă aparţinând unei metode, care, spre deosebire de experimentele psihofiziologice ale Şcolii din Leipzig, se caracterizează prin dimensiuni reduse, provine de la Galton. Probele se referă la randament, dexteritate, activitate. El vrea să stabilească un rezultat cantitativ, dar nu dezvăluie condiţiile psihologice şi fiziologice care se ascund în spatele randamentului.

Acest aspect behaviorist, care domină şi astăzi metoda testului, a fost introdus de Galton. Vestite şi încă în funcţiune sunt câteva teste inventate de el. Tubul lui Galton este un aparat pentru examinarea vederii în adâncime şi un măsurător al timpului de reacţie. De asemenea, a creat o hartă pentru precizarea acuităţii vizuale. Pentru diferenţierea culorilor a folosit bucăţi speciale de stofă, iar pentru distingerea mirosurilor, sticle cu diferite substanţe.

Punerea în aplicare a unor astfel de probe se realizează în cadrul Laboratorului antropometric, creat în 1884 şi deschis la International Health Exhibition din Londra, apoi la South Kensington Museum, care a funcţionat timp de şase ani, cu scopul de a măsura diferenţele individuale.

Măsurătorile efectuate erau de o mare varietate, în principal antropometrice. Printre acestea se numără : mărimea craniului, lungimea braţului, forţa mâinii (cu ajutorul unui dinamometru), înălţimea în poziţia şezând şi în picioare, lungimea degetului mijlociu, greutatea, capacitatea vitală a plămânilor (cu ajutorul unui spirometru), acuitatea vizuală şi auditivă, cel mai înalt ton auditiv detectabil, viteza unei lovituri, timpul de reacţie la stimuli vizuali şi sonori. Galton prefera să raporteze acuităţile senzoriale la inteligenţă şi, din acest motiv, laboratorul lui antropometric poate fi văzut

ca un efort pentru măsurarea inteligenţei. În 1888, Galton înfiinţează un laborator similar la South Kensington Museum, unde se lucrează câţiva ani. Laboratorul avea următoarele destinaţii:

1) Satisfacerea dorinţei persoanelor de a fi corect măsurate în anumite privinţe, pentru a obţine o aşteptare ipotetică la defecte remediabile în dezvoltare. Măsurătorile erau înregistrate la fiecare persoană. Se putea obţine o copie, sub restricţii rezonabile. Iniţialele şi datele de naştere erau introduse în registru, dar nu şi numele, care era indexat într-un registru separat.
2) Constituirea unui rezervor de informaţii pentru metodele practicate în domeniul măsurătorilor umane (este prea pretenţios să vorbim despre o bază de date).
3) Efectuarea de experimente antropometrice, de cercetări şi obţinerea de date pentru discuţii statistice.

Subiecţii erau primiţi de un asistent, care efectua şi măsurătorile. Pentru 3 pence (sumă modică la acea vreme) o persoană putea să fie măsurată în toate privinţele. Pentru un preţ mai mic (2 pence) o persoană putea să se măsoare din nou la o altă dată. Fiecare participant primea o copie a rezultatului sau a rezultatelor sale şi Galton păstra o copie pentru sine.

Numărul măsurătorilor înregistrate a fost foarte mare, considerând dificultatea inerentă oricărui început. S-a înregistrat un total de 16.000 de date individuale, dintre care 7.500 s-au păstrat până astăzi. Numai în primul an de funcţionare, Galton măsoară 9.337 de oameni.

El prezintă, de asemenea, probe care seamănă cu testele psihologice. Acestea sunt *aproape* teste. Li se acordă importanţă aspectelor teoretice ale trăsăturilor psihologice. În *Inquiries into Human Faculty and its Development*, face un bilanţ al studiilor asupra variabilităţii şi caracterului ereditar al funcţiilor mintale, în care prezintă diverse proceduri destinate să măsoare acuitatea senzorială, capacităţile de discriminare senzorială sau vivacitatea imaginilor mintale. Din 1884, contribuţiile sale vor fi printre cele mai semnificative în domeniul cercetărilor diferenţiale din acele vremuri. Măsurarea timpului de reacţie este direct inspirată din tehnicile lui Wundt, pe care Cattell le-a prezentat lui Galton în timpul unui sejur la Londra, determinant în orientarea ulterioară a lucrărilor lui Cattell.

Mai ales încercările simpliste de a măsura intelectul nu au fost roditoare. Cu toate acestea, de la el a venit un impuls deosebit pentru mişcarea testelor psihologice, în vederea demonstrării faptului că pot fi create teste obiective şi că scoruri semnificative pot fi obţinute prin proceduri standardizate. Ceea ce este caracteristic la Galton este lărgimea schemei de cercetare, în fapt o adaptare darwinistă la problema variabilităţii în general.

Galton consideră că măsurarea acuităţii senzoriale contribuie la măsurarea inteligenţei, deoarece fiinţa umană cunoaşte lumea doar prin intermediul senzaţiilor. Cu cât sunt mai acute simţurile, cu atât o persoană se presupune că este mai inteligentă. Acuitatea senzorială constituie o funcţie a înzestrării naturale, iar Galton consideră că inteligenţa este moştenită în familie.

Eugenia. Concluziile lui Galton privind moştenirea inteligenţei îl conduc spre ideea reproducerii selective. Dacă inteligenţa este moştenită, atunci inteligenţa poporului nu poate fi îmbunătăţită decât încurajând naşterea celor străluciţi şi descurajând naşterea

celor mediocri. Răspunsul lui Galton la această problemă este *da*. El era adeptul ameliorării calităţii intelectuale a fiinţelor vii prin selecţia eugenică. În anul 1865, se pronunţă pentru selecţia ştiinţifică a cuplurilor şi obligarea guvernului să plătească posesia trăsăturilor dezirabile. De asemenea, consideră că guvernul trebuie să facă diferite cheltuieli educaţionale pentru fiecare persoană. După lectura lucrării *Hereditary Genius*, Darwin îi scrie vărului său şi-l anunţă că nu aderă la astfel de convingeri.

Controversa moştenit - dobândit (natură - educaţie). Nativismul extrem al lui Galton nu a rămas fără urmări. Savantul englez măsoară frecvenţa eminenţei printre moştenitorii părinţilor iluştri şi compară frecvenţa eminenţei printre moştenitori la populaţia generală. Pentru a realiza această comparaţie, Galton studiază moştenitorii judecătorilor, oamenilor de stat, comandanţilor militari, oamenilor de litere, oamenilor de ştiinţă, muzicienilor, pictorilor şi poeţilor. Rezultatele le publică în lucrarea apărută în 1869, *Hereditary Genius: An Inquiry into its Laws and Consequences*. Concluzia este că moştenitorii individualităţilor ilustre au mai multe şanse să fie la rândul lor iluştri, decât urmaşii acelor individualităţi care nu au această calitate. Galton observă că zelul şi vigoarea trebuie să fie asociate cu capacităţile moştenite, înainte ca eminenţa să poată fi atinsă.

În epocă, au existat reacţii la teoriile lui Galton. De pildă, Alphonse de Condolle (1806-1893) a remarcat în mod deosebit importanţa ambianţei în apariţia oamenilor de ştiinţă. Condolle apreciază că toleranţa, climatul religios, guvernământul democratic şi o economie prosperă au fost mai importante, în ceea ce priveşte apariţia oamenilor de ştiinţă, decât capacităţile moştenite.

Galton răspunde la aceste aserţiuni în următoarea sa carte, *English Men of Science: Their Nature and Nurture* (1874). Pentru a strânge informaţiile necesare lucrării, Galton a trimis un chestionar unui număr de 200 de oameni de ştiinţă de la Royal Society. Se înregistrează astfel prima aplicare a unui chestionar în psihologie. Oamenilor de ştiinţă le-au fost adresate întrebări factuale cu referiri pornind de la condiţiile lor politice şi religioase până la mărimea pălăriei. Suplimentar, aceştia erau rugaţi să explice interesul lor pentru ştiinţă şi motivaţia alegerii domeniului de activitate ştiinţifică. În final, erau întrebaţi dacă interesul lor pentru ştiinţă era înnăscut.

Galton a notat că un număr disproporţionat de mare dintre oamenii de ştiinţă erau scoţieni şi că aceştia lăudau sistemul liberal de educaţie şcolară şi completă. Invers, oamenii de ştiinţă englezi erau lipsiţi de amabilitate, lucru care se poate spune şi despre sistemul educaţional englez. Pe baza acestei descoperiri, Galton solicita ca şcoala engleză să fie reformată, pentru a o depăşi pe cea scoţiană. Aici, Galton recunoaşte importanţa ambianţei. Poziţia sa este revizuită, susţinând acum că potenţialul este moştenit, dar că el trebuie să fie dezvoltat de propria cultură. În 1874, Galton stabilea clar controversa potenţial natural - ambianţă (educaţie), care constituie focusul atenţiei în psihologia modernă.

În următoarea carte, *Inquiries into Human Faculty and Its Development* (1883), sprijină teoria nativismului bazal prin studiul gemenilor. El găseşte gemenii monozigoţi ca fiind foarte asemănători unul altuia, de fiecare dată când cresc foarte departe unul de altul, şi găseşte dizigoţii ca fiind disimilari, chiar atunci când sunt crescuţi împreună. Cu aceasta, iniţiativa studierii gemenilor devine foarte populară, scopul final fiind acela de a determina influenţa relativă a educaţiei în privinţa diverselor atribute, cum ar fi, de pildă, inteligenţa. Studiul gemenilor rămâne de actualitate şi în zilele noastre.

Testul de asociere a cuvintelor. În *Inquiries...* se oferă primul test de asociere de cuvinte. Galton scrie 75 de cuvinte, fiecare pe o foaie separată de hârtie. Când aruncă o privire la fiecare cuvânt, oferă imediat şi răspunsul la acesta pe o altă foaie de hârtie. El prezintă fiecare dintre cele 75 de cuvinte în patru variante diferite, randomizând cuvintele de fiecare dată. Galton remarcă trei aspecte la acest studiu:

1) răspunsurile la stimulii-cuvinte tind să fie constante; el a dat foarte adesea acelaşi răspuns la un cuvânt în toate cele patru tipuri de experienţă;
2) răspunsul său era, adesea, extras din experienţa copilăriei sale;
3) o astfel de procedură evidenţiază aspecte ale psihicului, niciodată revelate înainte.

Nu este cunoscută influenţa pe care Galton a avut-o asupra lui Carl Gustav Jung, însă munca lui Galton cu asociaţiile de cuvinte anticipează două teme de interes ale psihologiei: folosirea asociaţiei libere şi recunoaşterea motivaţiei inconştiente.

Demne de remarcat sunt şi studiile sale asupra a ceea ce se numeşte *mental imagery*. Galton a fost printre primii cercetători, dacă nu chiar primul, în studiul imaginaţiei. În *Inquiries...* raportează rezultatele răspunsurile subiecţilor chestionaţi la imagini în care anumite persoane sunt prezentate la un mic dejun. El consideră că abilităţile imaginaţiei erau distribuite în mod normal, cu câţiva indivizi total incapabili de imaginaţie şi alţii având această abilitate foarte dezvoltată. Consideră că unii dintre prietenii săi nu au abilitatea formei imaginilor. Dacă senzaţiile şi derivatele lor constituiau materialul întregii gândiri, aşa cum îşi asumă această teorie oamenii de ştiinţă, atunci de ce unii oameni de ştiinţă nu au aptitudinea de a forma şi de a utiliza imagini? Galton consideră că orice persoană care are imaginaţie posedă o asemenea abilitate.

În examinare, Galton era interesat şi de relaţiile test – retest, de diferenţele generate de variatele măsurători, de relaţiile diverselor măsurători: statusul socioeconomic, asemănarea dintre familii prin diferite măsurători etc. Numărul foarte mare de date înregistrate de Galton a făcut ca unele dintre ele să rămână neanalizate. Recent, Johnson, McClearn, Yuen, Nagoshi, Ahern şi Cole (1985) au clasificat analizele moştenite de la Galton în rezultate analizate chiar de Galton şi date care nu au fost atent analizate de acesta.

Conţinutul probelor lui Galton este determinat prin concepţia psihologică dominantă în epocă, asociaţionismul. Toate fenomenele psihologice, inclusiv cele mai complexe, rezultă din asocierea senzaţiilor elementare. Există puţine diferenţe, din acest punct de vedere, între psihologia engleză şi cea germană. Referitor la formă, se poate spune că probele lui Galton sunt mult mai uşoare, mai simple, mai uşor de aplicat decât dispozitivele experimentale ale lui Wundt. Observarea mai multor indivizi este o consecinţă a opticii diferenţiale, dar se îngrijeşte destul de puţin de aplicaţiile imediate ale testelor în domeniul sănătăţii, al muncii şi al educaţiei.

Tendinţa lui Francis Galton a fost aceea de a căuta caracteristici, de a le face măsurabile, de a colecţiona date într-un număr cât mai mare, atât fizice, cât şi psihice, şi de a le prelucra. El găseşte metodele pentru a evalua cantitativ observaţiile, sprijinindu-se pe statistică, în vederea stabilirii distribuţiei aptitudinilor şi a raporturilor acestora. Era obsedat de măsură, fiind convins că orice este măsurabil. El considera că, dacă nu se poate măsura toată naţiunea, atunci este bine să se măsoare cât mai mulţi indivizi. A excelat în măsurarea intelectului, prin media timpului de reacţie, şi în cercetarea discriminării senzoriale. Unele dintre cercetările sale în vederea măsurării unor fenomene par să ţină de domeniul bizareriilor. Exemplificăm câteva dintre aceste măsurători: determinarea

eficienţei rugăciunii (pe care o găseşte ineficace); determinarea prin măsurători a ţării cu cele mai frumoase femei; determinarea gradului de plictiseală al lecturilor ştiinţifice; determinarea lungimii frânghiei spânzuratului pentru ca acesta să nu fie decapitat în timpul sinistrei operaţii; determinarea locului în care trebuie tăiată o prăjitură, în aşa fel încât aceasta să nu se rupă.

Toate tehnicile sale de măsurare au fost catalogate în detaliu de către K. Pearson.

Dezvoltarea statisticii. Şcoala engleză de psihologie este reputată în calitate de şcoală psihometrică, având mari contribuţii în statistică şi teoria măsurării. Statistica – văzută ca ansamblu de metode care vizează culegerea, gruparea, prelucrarea, analiza şi interpretarea datelor – îşi face apariţia în ştiinţele sociale prin Adolphe Quételet (1796-1874), astronom, matematician şi statistician belgian, care este o bună perioadă de timp (după 1846) citat drept „cea mai mare autoritate în domeniu" (*Lettres sur la théorie de probabilité apliquée aux sciences morales et politiques*). A aplicat teoria probabilităţii la studiul fenomenelor sociale şi a introdus în circulaţia ştiinţifică noţiunea de om mediu, foarte controversată de-a lungul timpului. El observă că „legea deviaţiei de la medie" a lui Gauss şi Laplace este valabilă pentru biologie, prin conformarea unei bune părţi din măsurătorile sale la exigenţele curbei distribuţiei în formă de clopot.

Primul savant care a aplicat sistematic statistica în psihologie a fost Francis Galton, care s-a ocupat, cum am mai spus, de o mulţime de lucruri (unele foarte bizare), dar mai ales de inteligenţă şi aptitudini. El demonstrează că distribuţia normală, fundamentată teoretic de Gauss şi Laplace, este adecvată pentru mărimi ca greutatea, înălţimea, dimensiunea bustului. Nu este interesat doar de mărimile absolute, ci şi de stabilirea de poziţii pe o scală de evaluare pe baza căreia se pot acorda note. Practic, în această realizare se obiectivează nevoia unei scale în care teoria şi faptele pot să concorde mai bine. Distribuţia normală are meritul de a clasifica diferite grade ale inteligenţei, iar Galton a formulat unul dintre cele mai importante principii ale măsurării viitoare a inteligenţei, şi anume acela conform căruia persoanele supradotate se găsesc peste medie în aceeaşi proporţie în care cele subdotate se află sub medie.

Galton consideră că *legea deviaţiei de la medie* poate fi utilizată în scopul de a cuprinde gradientul relativ de consonanţă între două variabile. Cu aceasta introduce tehnica corelaţiei în biologie şi psihologie, dezvoltând un procedeu de corelaţie simplu, pe care l-a numit *indice de corelaţie*. În 1888, Galton a publicat un studiu intitulat *Co-relation and Their Measurement, Chiefly from Anthropometric Data*, iar în 1889, *Natural Inheritance*. Ambele lucrări descriu conceptele de *corelaţie* şi *regresie*.

Galton definea corelaţia sau *co-relaţia*: „Două organe variabile se spune că se află în corelaţie când variaţia unuia este însoţită, în medie, de mai mici sau mai mari variaţii ale celuilalt şi în aceeaşi direcţie. Astfel, lungimea braţelor este corelată cu cea a picioarelor, deoarece o persoană cu braţele lungi are, de regulă, picioare lungi şi invers" (*apud* Hergenhahn, 1997, p. 275).

În plus, la introducerea conceptului de *corelaţie*, Galton indică mediana ca o măsură mai adecvată pentru evidenţierea tendinţei centrale. El găseşte media prea influenţată de scorurile extreme ale distribuţiei şi preferă să utilizeze *scorul mediu cel mai bun* al unei distribuţii (mediana într-o distribuţie).

Edge Worth introduce noţiunea de *coeficient de corelaţie*, iar K. Pearson dezvoltă, prin utilizarea datelor cuprinse în încercările lui Bravais, forma finală a corelaţiei prin momentul produselor. K. Pearson, F. Galton şi W.F.R. Weldon au fondat în anul 1900

revista *Biometrika. A Journal for the Statistical Study of Biological Problems*, destinată cercetării matematice în biologie şi psihologie. Cu aceste contribuţii se poate considera că s-a intrat în faza sistematică a cercetărilor cu caracter psihologic.

O figură de primă mărime a statisticii engleze şi mondiale, cu mari realizări în domeniul de aplicaţie psihosocial, este Karl Pearson (1857-1936), întemeietorul şcolii anglo-saxone de statistică matematică. Corelaţiile, statistica χ^2, teoria erorilor, momentele şi teoria repartiţiilor sunt câteva dintre preocupările sale de vârf. Ca o recunoaştere a contribuţiilor lui K. Pearson în analiza datelor experimentale, acestuia i se rezervă o catedră de cercetare în cadrul Laboratorului de biometrie al Universităţii din Londra, o dată cu fondarea sa, în anul 1904. Ulterior, a editat şi *Annals of Eugenics*, publicaţie a Laboratorului de Eugenie. Karl Pearson a fost profesor şi la celebra Universitate din Cambridge.

O altă personalitate de prim rang a ştiinţei engleze, de la sfârşitul secolului al XIX-lea şi începutul secolului XX, cu preocupări în domeniul statisticii cu aplicaţii în psihologie, a fost Charles Spearman (1863-1945). Dacă Pearson a fost în primul rând matematician şi statistician care a aplicat cunoştinţele sale în domenii psihosociale, Ch. Spearman a fost înainte de toate psiholog de formaţie, cu o pasiune deosebită pentru statistică, numele lui înscriindu-se definitoriu în istoria acestei discipline. Evoluţia sa profesională este deosebit de curioasă. După o carieră militară în armata engleză, până la vârsta de 34 de ani, Ch. Spearman îşi îndreaptă interesul spre psihologie, studiind cu W. Wundt şi O. Külpe în Germania. Îşi întrerupe studiile pentru a sluji în armata engleză în timpul războiului burilor (1900-1902), perioadă în care citeşte opera lui Galton. Reuşeşte performanţa unei serii de experimente în rândul şcolarilor şi rezultatele par să confirme ideile lui Galton, referitoare la relaţiile dintre acuitatea senzorială şi inteligenţă. El descoperă importanţa măsurătorilor asupra acuităţii senzoriale care se corelează puternic, dar şi pe cea a corelării înalte între acestea şi „deşteptăciunea elevilor în şcoală" (+0,38) (vezi Hergenhahn, 1997, p. 282). În urma publicării studiilor sale, controversate de altminteri, privind determinarea şi măsurarea inteligenţei generale, lui Spearman i s-a oferit un post la University College din Londra, unde începe acţiunile de combatere a senzorialismului, asociaţionismului, hedonismului şi a multor alte concepte filosofice şi psihologice acceptate. Atacurile sale erau rezultatul studiilor, precum cele ale lui Wissler, care găseau mici intercorelaţiile între măsurătorile lui Galton şi Cattell asupra acuităţii (în cele mai multe cazuri, nu existau corelaţii între măsurătorile acuităţilor senzoriale şi performanţele academice). Astfel, în anul 1904, Charles Spearman publică două lucrări de statistică[1], după ce, în 1901, Clark Wissler prezentase lucrări de corelaţie statistică asupra rezultatelor testelor lui J. McKeen Cattell. Într-o primă lucrare, Spearman demonstrează importanţa tehnicilor de corelaţie statistică pentru scopuri practice de cercetare, iar în a doua, pune fundamentul pentru tehnica analizei factoriale, care încununează statistica corelaţiei, fapte ce prezintă mare importanţă pentru cercetarea testelor de inteligenţă generală şi a funcţiilor specifice ale acesteia. El schiţează, totodată, un model al ierarhiei inteligenţelor. În 1901, împreună cu Krueger, realizează diferite corelaţii între diversele niveluri ale aptitudinilor psihice.

1. Cele două studii se intitulează „The proof of measurements of association between two things", in *American Journal of Psychology*, nr. 15, pp. 72-101, şi „General intelligence, objectively determined and measured", in *ibidem*, pp. 201-292.

Pentru că rezultatele lui erau în opoziţie, de cele mai multe ori, cu ceea ce se scrisese până atunci în domeniul său de interes, el considera că rezultatele contrare propriilor rezultate erau artefacte. Spearman ajunge la concluzia că, din cauza intercorelării măsurătorilor acuităţii senzoriale, acestea trebuie să măsoare o abilitate sau o facultate comună pe care o numea inteligenţă generală (g). Urmând tradiţia lui Galton, el susţinea că *g*-ul era determinat, de cele mai multe ori, exclusiv prin zestrea genetică. Se pot face câteva observaţii privind concepţia lui Spearman asupra inteligenţei:

- inteligenţa este o facultate unitară;
- inteligenţa este, în mare parte, moştenită;
- în mare măsură, concepţia lui Spearman despre inteligenţă era racordată la noua mişcare a testelor din Statele Unite, şi nu la concepţia lui Binet.

2.3. Contribuţia şcolii psihologice franceze la apariţia testului psihologic

Psihologia franceză are, prin excelenţă, preocupări ştiinţifice în privinţa psihopatologiei. Curentul psihomedical se conturează încă de la Descartes, propagându-se în secolele următoare prin reprezentanţi străluciţi, unii dintre ei amintiţi anterior. Prin analiza pe care o permite, psihopatologia poate fi considerată un substitut al metodei experimentale, dar, în esenţă, ea este o psihologie clinică. Deşi se interesează de diferenţele individuale, metoda clinică nu incită în nici un fel la măsurarea acestora. O dată cu Théodule Ribot (1839-1916), această preocupare devine autonomă în cadrul curentului medical. Opera lui va fi continuată de Pierre Janet (1859-1947) şi de G. Dumas (1866-1946). Ribot susţine că degradarea psihopatologică marchează procesele mintale în ordinea inversă a apariţiei lor. Ea face posibilă disocierea proceselor care sunt, de obicei, bine integrate în funcţionarea psihologică normală. Studiul psihopatologiei reprezintă un mijloc pentru mai buna înţelegere a funcţionării psihologice normale. Alfred Binet, căruia îi vom acorda în această lucrare un spaţiu relativ extins, poate fi considerat ca aparţinând şcolii franceze de psihopatologie. Sub influenţa lui Ribot, el se orientează spre psihologie şi, la sfaturile lui Charcot, publică primele articole de psihopatologie. Apartenenţa sa la această şcoală explică interesul pentru fenomenele mintale globale şi atitudinea de respingere a atomismului psihologic.

Având oarecum rădăcini în şcoala franceză, psihologia americană anterioară apariţiei testului psihologic a contribuit mai puţin la fundamentarea teoretică a apariţiei respectivului tip de test. Se consideră că James McKeen Cattell are o puternică amprentă europeană în formaţia sa ştiinţifică. Situaţia contribuţiei psihologilor americani la evoluţia testului psihologic se va schimba, mai ales după apariţia scalei Binet-Simon. William James (1842-1910) este considerat în mod unanim fondatorul psihologiei americane. Acesta a avut o formaţie medicală şi a suferit influenţa psihologiei germane, ca majoritatea primilor psihologi americani. Destul de repede se distanţează faţă de influenţa psihologiei experimentale, pentru a se consacra psihologiei axate pe descrierea stărilor de conştiinţă, acordând un loc aparte reflecţiei filosofice şi sociale, iar în cadrul cercetărilor psihologice, spiritismului şi telepatiei. Prin îndepărtarea de Wundt, el favorizează naşterea psihologiei funcţionaliste şi pragmatice care nu poate decât să faciliteze apariţia testelor. Psihologii care urmează lui James, Cattell şi, mai ales, Stanley Hall, creează în 1892

American Psychological Association şi sunt partizani ai funcţionalismului înclinat spre aplicaţii. Funcţionalismul are drept obiect funcţiile conştiinţei, şi nu structurile acesteia. El este foarte legat de teoria evoluţiei, o importanţă deosebită având-o circumstanţele în selecţia şi „punerea în operă" a funcţiilor psihice, ceea ce oferă psihologiei perspective practice. Psihologii funcţionalişti au abandonat introspecţia, dar au integrat spiritul experimental din psihologia germană şi preocupările pentru măsură din psihologia engleză. Construcţia testelor va fi una dintre marile lor probleme şi, datorită promovării testului, ei vor reuşi să constituie un grup profesional puternic.

2.4. Contribuţia psihiatrilor la apariţia testului psihologic

O cale relativ autonomă în pregătirea apariţiei testului psihologic urmează psihiatrii germani C. Rieger (1888), R. Sommer (1896), E. Kraepelin (1896) A. Oehrn (1889), Th. Ziehen (1897/1909) şi A. Starring, care au căutat să dezvolte metode mai cuprinzătoare ale diagnosticului psihiatric[1]. Examinarea capacităţii mintale cu ajutorul probelor psihologice, în sensul de situaţii problematice nesistematizate, ia avânt în rândul psihiatrilor începând de la jumătatea secolului al XIX-lea. Prezenţa acestora este relativ numeroasă, dar absenţa standardizării şi deficienţele de construcţie fac ca ele să fie date uitării. Totuşi, creaţiile respective au influenţat cursul testării psihologice. Aplicaţii tipice pentru astfel de „teste" a făcut fizicianul german Hubert von Grashey, examinând cu ajutorul lor memoria persoanelor care prezintă răni la cap. O astfel de probă prezintă, printr-o fantă, cuvinte, tablouri, simboluri desenate pe o foaie de hârtie mişcată încet deasupra ochilor bolnavului. Respectivul model de solicitare este preluat de multe teste psihologice relativ recente. În 1889, psihiatrul german Conrad Rieger elaborează un proiect ambiţios de a determina defectele în urma traumatismului cranian. Bateria de probe este compusă din câteva secţiuni, incluzând recunoaşterea simbolurilor sociale, memoria de scurtă durată prin stimulare olfactivă, memoria de lungă durată pentru materiale învăţate anterior, perechi de asociaţii (de exemplu „Sodoma şi Gomora"), alcătuirea unor propoziţii cu un număr redus de cuvinte date pacientului, imitări ale mişcărilor şi propoziţiilor vorbite. Bateria lui Rieger era nepractică, întrucât presupunea mai mult de 100 de ore în administrare. Totuşi, această baterie va influenţa subtestele de formare a propoziţiilor prezente în câteva dintre primele teste de inteligenţă. Anumite proceduri, standardizate ulterior, au fost perpetuate până în zilele noastre. C. Rieger este primul în Germania (alături de Galton, în Anglia) care a încercat să dezvolte o metodă de diagnoză a inteligenţei, cu un drum cazuistic ce cuprinde analiza multilaterală a inteligenţei. El încearcă un diagnostic diferenţiat, sarcină impusă de practica clinică, şi îşi demonstrează metoda prin intermediul cazurilor clinice de tulburări ale inteligenţei. Diferenţele specifice se stabilesc la un individ valid anterior. Prin metode psihologice şi psihofiziologice, el examinează ceea ce consideră a fi inteligenţa şi ipotezele sale elementare: percepţia, apercepţia, uitarea, procesele asociative, recunoaşterile etc.

Fiecare domeniu era cercetat după indicaţii metodice detaliate. Ceea ce lipseşte este comparabilitatea rezultatelor individuale prin procedee statistice care ar da indicaţii despre un anumit loc într-o ierarhie valorică (W. Stern, 1911). Pentru fundamentarea normalităţii se expun tulburări ale inteligenţei, clinic relevante.

1. Anul dintre paranteze marchează apariţia unor lucrări pe tema care ne interesează.

Diagnosticul psihopatologic a mers mult timp pe drumul trasat de Rieger.

În anul 1897, Th. Ziehen reduce substanţial schema cuprinzătoare de cercetare, oprindu-se în final la memorie, abstracţie, activitatea combinatorică. Eforturile metodice ale lui Ziehen constau într-o serie de sarcini în formă nestandardizată, care sunt utilizate astăzi ca teste. Şi la el lipseşte baza de raportare statistică. În 1911, W. Stern realizează o critică a cercetării psihopatologice din vremurile de început care se bazau pe rezultatele cercetărilor empirice ale lui E. Rodenwaldt.

Un alt grup de cercetători au pus în prim-plan interesul pedagogic; printre aceştia îi putem enumera pe J. McKeen Cattell şi L. Farrand (1896), H. Ebbinghaus (1897), E.L. Thorndike (1903) şi E. Meumann (1905).

James McKeen Cattell (1860-1936) crează, în 1890, noţiunea de *mental test* în lucrarea sa *Mental Test and Measurement*. Cel ce avea să „boteze" probele psihologice cu numele de teste a studiat în Germania şi Anglia, care, pe atunci, deţineau clar primatul cercetărilor experimentale în psihologie. În Germania, din 1880, studiază în primul rând cu fiziologul kantian R.H. Lotze. În anul 1882, se întoarce în SUA, scrie o carte de filosofie şi studiază la Johns Hopkins University. În această perioadă face cercetări de psihologie în laboratorul lui G. Stanley Hall, nutrind dorinţa de a deveni psiholog. Revenind în Germania, în 1883, Wilhelm Wundt îl acceptă să lucreze în laboratorul de psihologie experimentală creat de el în anul 1879. Este primul asistent experimental şi primul student din Statele Unite care a urmat un doctorat supervizat de W. Wundt. Diploma corespunzătoare o primeşte în anul 1886. Împreună cu un coleg, James McKeen Cattell face numeroase studii asupra timpului de reacţie, cercetări în vogă în perioada de început a psihologiei experimentale. Influenţa lui Wundt este prezentă prin măsurările pe care le efectuează: timpul de reacţie cerut de diferite reacţii mintale, măsurat în fracţiuni de secundă. Deoarece rezultatele sale difereau de cele ale unui coleg al său, îi propune lui Wundt ca diferenţele individuale în privinţa timpului de reacţie să fie exploatate ştiinţific, propunere respinsă deoarece savantul german era mult mai interesat de natura generală a psihismului decât de diferenţele individuale. După obţinerea doctoratului şi revenirea în Statele Unite, începe o corespondenţă ştiinţifică cu Francis Galton, tema predilectă fiind timpul de reacţie. În următorii doi ani, J. McKeen Cattell realizează cercetări la Cambridge University, unde lucrează cu celebrul om de ştiinţă englez. Cattell manifestă un interes deosebit pentru diferenţele individuale, iar Galton îi confirmă convingerile conform cărora diferenţele individuale erau importante şi puteau fi obiectiv măsurate. Sub influenţa lui Galton, Cattell ajunge să creadă că inteligenţa se raportează la acuitatea senzorială şi, în consecinţă, aceasta reprezintă inteligenţa copiilor. Se consideră discipol al lui Galton şi în privinţa eugeniei. Ca atare, propune ca persoanele dotate să beneficieze de consideraţia datorată celor mai bune elemente ale unui popor. De asemenea, el crede potrivit ca persoanele respective să se căsătorească cu indivizi similari, pentru a avea familii numeroase. El promite celor şapte copii ai săi stimulente în cazul căsătoriei lor cu copii ai colegilor profesori.

După ce s-a întors în Statele Unite, Cattell a fondat la University of Pennsylvania, în anul 1890, un laborator de psihologie destinat studenţilor şi a dezvoltat o serie de teste care au fost, în fapt, extensii şi ajustări la bateriile lui Galton. Sunt probe simple care

C:da 239 coala 3

progresează de la măsurările corporale la cele predominant psihice. James McKeen Cattell recunoaşte că unele teste sunt luate din laboratorul lui Galton, pe care le-a acceptat ca teste pur corporale. Seria constă din măsurători ale forţei de presiune cu ajutorul dinamometrului, ale vitezei de mişcare, dar şi cu privire la sensibilitatea pielii, sensibilitatea la durere, sensibilitatea în aprecierea diferitelor greutăţi, la timpul de reacţie la stimuli acustici, timpul necesar pentru denumirea culorilor, cu referire la împărţirea unei linii de 50 cm în două părţi, estimarea a 10 secunde, durata reproducerii unor litere la o singură auzire. Se dau indicaţii asupra respectării instrucţiunilor de cercetare. După aceste încercări, nu oferă impresia că posedă o reprezentare diferenţiată a inteligenţei şi a aptitudinii de randament intelectual. Conştiinţa şi inteligenţa nu vor fi clar separate una de cealaltă. Cattell afirmă că nu s-ar putea preciza în mod limpede ce valori de simptom au unele teste. De asemenea, realizează, practic, transferul procedurii de testare pe care a studiat-o cu Wundt şi Galton pe tărâm american.

În studiul în care folosea noţiunea de *mental test*, Cattell spune: „Psihologia nu poate pretinde certitudinea şi exactitatea ştiinţelor fizice, dacă nu se fondează pe experiment şi măsurare. Un pas în această direcţie poate fi făcut aplicând o serie de *teste mintale* şi măsurători la un mare număr de indivizi. Rezultatele vor fi de considerabilă valoare ştiinţifică în descoperirea constanţei proceselor mintale, interdependenţei şi variaţiei lor în diferite circumstanţe" (*apud* Hergenhahn, 1997).

În 1891, Cattell se transferă la Columbia University, cea mai mare universitate din Statele Unite la acea vreme, unde timp de 26 de ani ocupă postul de profesor, fiind fără îndoială decanul psihologilor americani. Reputaţia sa a adus numeroşi studenţi printre care îi enumerăm pe E.L. Thorndike, R.S. Woodworth, E.K. Strong şi C. Wissler. Aici începe să administreze testele sale unor studenţi debutanţi. În programul de testare al lui Cattell exista prezumţia că, dacă testele sale măsurau acelaşi lucru (inteligenţa), performanţele la astfel de teste trebuiau să fie înalt corelate. Dacă un test de inteligenţă este valid, cu ajutorul rezultatelor acestuia se pot face predicţii diferenţiate asupra acelor performanţe individuale care pretind inteligenţă în timpul realizării lor.

În 1901, Clark Wissler, unul dintre studenţii *graduate* ai lui Cattell, testează prezumţiile sale. El obţine scorurile testelor şi notaţiile academice de la peste 300 de studenţi de la Columbia University şi Barnard College. Scopul aplicaţiilor era de a afla dacă scorul testelor poate prevedea performanţele academice, urmărindu-se practic un scop de validare. Folosind coeficientul de corelaţie al lui Pearson, Wissler măsoară relaţiile dintre rezultatele obţinute de studenţi la testele lui Cattell şi performanţele academice ale celor testaţi psihologic. Rezultatele obţinute de Clark Wissler au fost descurajante pentru programul de testare elaborat de Cattell. El găseşte că intercorelaţiile între teste erau foarte slabe, iar cele dintre diversele teste şi succesul în mediul academic erau foarte apropiate de 0. Analiza corelaţiilor cu performanţele academice indică nişte corelaţii modeste: 0,16 cu memoria, 0,08 cu dinamometrul, 0,02 cu denumirea culorilor. Verificarea *instrumentelor de alamă* arată în mod clar o slabă predictibilitate. Între denumirea culorilor şi mişcarea mâinii se găseşte o corelaţie de 0,19, iar între timpul de reacţie şi denumirea culorilor 0,15.

Concluzia trasă din rezultatele respective era că testele nu măsoară aceleaşi lucruri, deoarece, dacă s-ar fi întâmplat aşa, ele ar fi trebuit să aibă o corelaţie înaltă, şi că ele nu erau valide, deoarece, în acest caz, scorurile ar fi trebuit să fie înalt corelate cu achiziţiile academice. După această constatare, interesul savanţilor pentru testele psihologice

a scăzut simţitor. Ca urmare a eşecului, Wissler îşi schimbă preocupările în domeniul antropologiei, domeniu în care îmbrăţişează concepţia ambientalistă în privinţa aptitudinilor umane, iar Cattell îşi îndreaptă atenţia spre alte aspecte ale psihologiei aplicate. Deja devine limpede faptul că probele galtoniene nu erau valide pentru a măsura inteligenţa. Este sfârşitul unei iluzii. Datorită eşecurilor înregistrate, psihologii experimentali ai acelor vremuri au abandonat folosirea timpului de reacţie şi discriminarea senzorială ca măsuri ale inteligenţei. Renunţarea a fost însă prematură şi pripită. Dacă folosim standardele de apreciere contemporane, putem afirma că metodele de cercetare ale lui Wissler relevă o extraordinară naivitate psihometrică. Faptul că el a folosit studenţi strălucitori a dus, inevitabil, la micşorarea corelaţiilor. Dacă grupurile ar fi fost mai eterogene, corelaţiile ar fi fost cu siguranţă mai mari. Colectarea seriilor de valori de genul timpului de reacţie a fost, în mod firesc, lipsită de fiabilitate, dat fiind numărul mic de încercări pe subiect. Un asemenea deficit plasează corelaţia în zona unor valori marginale. Ulterior, importanţa timpului de reacţie a fost reevaluată, în privinţa măsurării inteligenţei. Galton şi elevii săi aveau convingerea că din aptitudinile de diferenţiere senzorială se pot trage concluzii asupra capacităţii de decizie intelectuală. Golul creat de abandonarea tradiţiei galtoniene nu va dura mult şi americanii vor continua dezvoltarea testelor, virtual încetată la cotitura secolului al XIX-lea. Cercetările lui J. McKeen Cattell nu au fost singulare în Statele Unite. În 1892, Jastrow testează vizitatorii unei expoziţii şi compară rezultatele copiilor cu aprecierile profesorilor. De asemenea, Gilbert formează un eşantion sistematic de copii şi tineri între 6 şi 18 ani, după trepte de vârstă şi sex, şi compară rezultatele cu deciziile profesorilor. În 1898-1899, Sharp parcurge drumul lui Cattell şi demonstrează clar următoarele:

- *mental test* oferă o mică intercorelaţie;
- *testele fizice* indică o corelaţie generală unul cu altul;
- *notele-martor ale studenţilor* corelează într-o măsură considerabilă, dar nu cu testele de laborator.

În acest fel, s-a îngropat o mare speranţă care se lega de testele mintale ale lui Cattell şi Galton. Cattell dovedeşte că avea un potenţial de cercetare elementar, fapt ce i-a determinat pe unii psihologi din Europa să se dedice altor preocupări (Binet în Franţa, Ebbinghaus în Germania).

Deruta psihologilor în legătură cu lipsa de relevanţă a cercetărilor efectuate de Cattell nu a durat mult şi nu a însemnat o încetare a producţiei de teste psihologice sau abandonarea preocupărilor ştiinţifice pentru acestea.

Începând cu anul 1899, au fost publicate în Europa numeroase teste. A. Oehrn publică la Heidelberg teste de percepţie, de memorie, de asociaţie şi privitoare la funcţiile motrice. Elevul lui, Emil Kraepelin, propune utilizarea acestor teste în psihiatrie. În 1891, Hugo Münsterberg, care va emigra în SUA în anul următor, prezintă în Germania 15 teste, cu intenţia de a studia cum se diferenţiază indivizii după studiile şi profesiile lor.

Din aceste prime tentative de a construi teste, se poate observa că toate bateriile de probe propuse au în comun faptul că privilegiază observaţia proceselor elementare, senzoriale şi perceptive, şi a timpului de reacţie. Anumiţi psihologi încep să-şi pună întrebări asupra fundamentelor acestei alegeri, aşa cum fac Münsterberg şi Baldwin, care îşi propun să acorde memoriei mai multă importanţă, tendinţa respectivă fiind, în continuare, foarte puternică. Ea este evidentă în colectarea probelor (larg inspirată din

Wundt) prezentate în 1904 în lucrarea *Technique de psychologie expérimentale*, avându-i ca autori pe Toulouse, Vaschide şi Piéron. În a doua ediţie a acestei opere (1911), se consacră 250 de pagini fenomenelor senzoriale şi 102 fenomenelor intelectuale (62 pentru memorie, 22 pentru fenomene asociative, 13 pentru fenomene logice şi 5 pentru raţionament). Desigur, este surprinzătoare insistenţa asupra proceselor elementare, când se propune în primul rând abordarea fenomenelor mintale. Acest fapt este consecinţa alegerii postulatelor empiriste (asociaţionismul) şi a voinţei de a construi psihologia pe modelul ştiinţelor naturii. Dar probelor astfel elaborate le lipseşte credibilitatea pe plan teoretic, ceea ce a dus la un impas şi în plan practic. Pe plan teoretic, există o distanţă enormă între pretenţii şi observaţiile efectiv realizate. Henry Walon spunea în 1930: „Sau cercetările întreprinse rămân un pur exerciţiu de laborator, sau trebuia un joc de asimilări arbitrare pentru a umple intervalul care se deschidea între rezultatele lor foarte particulare şi o reprezentare încă grosieră şi factică a funcţiilor psihice. În loc de a oferi o semnificaţie prin ele însele, singura cale pe care ei au putut să o admită a fost de a le raporta la unele dintre entităţile create prin introspecţie. Astfel, psihologia care, pentru a fi ştiinţifică, se ataşează la tehnică nu ştie decât să adopte ideologia aceleia căreia îi repudiază cel mai mult metodele" (Huteau şi Lautrey, 1995, p. 19).

2.5. Comanda socială de psihodiagnostic al caracteristicilor individuale la graniţa dintre secole

La sfârşitul secolului al XIX-lea, ţările din Europa de Vest cunosc transformări culturale, sociale şi economice profunde, context în care se remarcă şi o puternică cerere de cunoştinţe psihologice operatorii. Această cerere se face simţită în numeroase sectoare, cum ar fi industria, comerţul, publicitatea, educaţia specializată, justiţia, sănătatea, politica etc. Ea corespunde nevoilor privind problemele de evaluare a persoanelor, de gestiune şi organizare, de formare. Psihologia aplicată se naşte, se dezvoltă şi se diversifică pentru a răspunde unor astfel de cerinţe, privind caracterizarea persoanelor în termeni de eficienţă cognitivă, fapt care va presupune punerea la punct a testelor psihologice, în primul rând pentru instituţiile de asistenţă a deficienţilor mintal, în şcoala primară şi în mediul industrial.

Cererea mijloacelor permiţând diagnosticarea eficienţei intelectuale apare, mai ales, în legătură cu dificultăţile mintale accentuate, adică idioţenia şi imbecilitatea. Ea nu corespunde doar unei voinţe de a distinge şi de a clasa, ci şi unui imperativ practic provenit din abandonul postulatului de incurabilitate a unor astfel de deficienţe şi, în consecinţă, din crearea şcolilor de asistenţă specializate. Am amintit anterior că prima şcoală de acest gen a fost fondată de Philippe Seguin, care după emigrarea în SUA (1848) va juca un rol foarte important în dezvoltarea psihopedagogiei speciale. În cele aproximativ 80 de şcoli de copii şi adulţi deficienţi mintal existente, în 1870, în SUA, se punea problema identificării veritabililor deficienţi mintal şi a diferenţierii lor de cei care perturbă sau jenează, acolo unde se găsesc. Exista tendinţa, în şcolile normale, de debarasare de acei elevi cu uşoare tulburări de comportament, dar care nu aveau deficienţe de ordin mintal. De asemenea, exista dorinţa de a crea clase relativ omogene. Din cele două motive sunt comandate de către diriguitorii din învăţământ metode permiţând evaluarea eficienţei şi dezvoltării cognitive.

Nevoi de acelaşi tip apar şi în sectorul judiciar. Atunci când deficienţele intelectuale nu permit distingerea între bine şi rău, nu există responsabilitate, lipsă care, o dată constatată, duce la absolvirea de condamnare juridică şi de pedepsire a autorilor unor delicte. Dar, pentru a constata neatingerea *vârstei raţiunii*, este nevoie de o expertiză psihologică presupunând, la rândul ei, instrumente de investigaţie adecvate. Ideea câştigă teren prin psihiatrul german Emil Kraepelin, elev al lui Wundt, care susţine cu argumente că practica psihiatrică va fi ameliorată dacă va integra tehnici obiective de caracterizare a persoanelor, mai ales în domeniul funcţiilor cognitive.

La sfârşitul secolului al XIX-lea, învăţământul primar se generalizează în majoritatea ţărilor industrializate şi se creează foarte repede condiţii juridice şi materiale pentru înfiinţarea unui învăţământ special, destinat copiilor care, datorită deficienţei lor intelectuale, nu au posibilitatea de a urma învăţământul normal. După câteva tentative nereuşite în Germania, primele clase pentru şcolari retardaţi sunt deschise în SUA şi, din 1904, sunt create comisii ministeriale în Franţa şi Anglia pentru a studia această delicată problemă. În Franţa, recomandările date de Comission Bourgeoise conduc la crearea unor clase de perfecţionare, în 1909. O perioadă apreciabilă de timp, s-a considerat că aceste clase speciale s-au constituit pentru a elimina jena copiilor din clasele normale incapabili să urmeze programele şcolare, fiind astfel excluşi de la şcolarizarea obişnuită. În fapt, crearea de clase speciale era mai degrabă o consecinţă a existenţei unui proiect umanist, iniţiat de Bourneville (1840-1909) şi susţinut de F. Boisson şi A. Binet, decât manifestarea unei voinţe de omogenizare a populaţiei şcolare. Crearea unor astfel de clase punea problema detectării copiilor deficienţi. După 1890, Binet se interesează de deficienţa mintală şi depune eforturi energice pentru crearea învăţământului special. Instrumentul de investigaţie creat de el are menirea de a ajuta la detectarea *anormalilor* din şcolile învăţământului de masă.

Revoluţia industrială din a doua jumătate a secolului al XIX-lea are drept consecinţe generalizarea, accelerarea şi diversificarea procesului de diviziune a muncii. Organizarea muncii după principiile lui Frederick Winslow Taylor duce la separarea strictă între concepţie şi execuţie, la parcelizarea scopurilor şi la sărăcirea muncii, ca preţ al eficienţei economice. Apar funcţii profesionale total diferite de cele existente până atunci. Procedurile tradiţionale de recrutare şi de afectare a mâinii de lucru sunt bulversate, fiind înlocuite cu noi metode de selecţie profesională. În afară de selecţie, apare necesitatea unui sistem nou de orientare profesională a tinerilor. După 1910, psihologii încep să elaboreze teste în scopul de a contribui la rezolvarea acestei probleme. În bună parte, psihologii erau antitaylorişti şi manifestau o grijă deosebită de a rămâne neutri în conflictele de muncă. Nu trebuie să vedem în toţi psihologii din această perioadă fiinţe interesate exclusiv de ştiinţă, unic motivaţi prin cunoştinţe şi întorşi de la interesele lor individuale pentru a face în aşa fel încât societatea să beneficieze de munca lor. O astfel de imagine ar fi idilică, lipsită de realism. Majoritatea dintre ei sunt orientaţi pragmatic, spre aplicaţii, deosebit de activi în promovarea propriilor metode. Sunt preocupaţi permanent de a face profitabilă din punct de vedere economic profesiunea lor. Ei nu stabilesc distincţia netă între psihologia fundamentală şi psihologia aplicată şi sunt convinşi că indivizii trebuie studiaţi în mediul în care sunt inseraţi, în legătură cu sarcinile ce le sunt prescrise în plan social.

3. Momentul Binet-Simon în evoluţia testului psihologic

Cu toate aceste evidente realizări (unele dintre probele psihologice concepute înainte de 1900 sunt utilizate şi astăzi, de exemplu testul Bourdon-Amfimov, creat în 1895), prototipul testului psihologic îl reprezintă scala de inteligenţă realizată de Alfred Binet şi Théodore Simon. Profesorul american Robert Gregory de la Wheaton College (Illinois) susţine, într-o bine cunoscută lucrare dedicată testelor psihologice (1996, p. 12), că scala metrică creată de Binet şi Simon se sprijină pe un test *a priori*, dezvoltat de doctorul american Blin şi elevul său Domaye, în anul 1902. Totuşi, trebuie amintit faptul că nu orice situaţie problematică expusă într-o anumită lucrare poate fi considerată test fără o minimă verificare. Astfel de realizări pot fi apreciate în aceeaşi măsură ca şi situaţiile problematice de care sunt pline basmele pentru copii.

Există o similitudine în destinul a doi mari creatori de produse specifice civilizaţiei moderne: Henry Ford şi Alfred Binet. Primul este considerat întemeietorul erei automobilului, cu toate că, înainte de a lansa Ford T pe piaţă, au fost construite zeci de modele de vehicule mici destinate transportului personal. Recunoaşterea ca părinte al automobilului, de care se bucură Ford, se datorează faptului că acesta a reuşit să influenţeze astfel benefic viaţa practică a societăţii şi a reuşit să facă acest mijloc de transport accesibil unei mari mase de oameni, prin ideile sale geniale de ordin tehnologic, organizatoric şi financiar.

Alfred Binet este considerat părintele metodei testului şi al instrumentului de investigaţie purtând numele de *test psihologic*, cu toate că înaintea elaborărilor sale au circulat numeroase probe cu pretenţia argumentată de a se intitula astfel. Chiar denumirea de *test* a fost dată de americanul James McKeen Cattell cu 15 ani înaintea apariţiei scalei metrice a inteligenţei. Prin perfecţionarea metodologiei de construcţie şi prin concepţia realistă privind menirea testului psihologic, Alfred Binet a reuşit să introducă testul în viaţa instituţiilor sociale şi să-l facă util pentru un mare număr de oameni. Practic, prin creaţia lui Binet, testul psihologic a primit botezul social[1].

Alfred Binet (1857-1911) s-a născut la Nisa, într-o familie de intelectuali cu preocupări în domeniul fizicii (tatăl şi ambii bunici erau fizicieni). El nu continuă tradiţia familiei, urmând în schimb studii medicale. Lecturile preferate sunt lucrările lui Darwin, Galton şi cele ale empiriştilor englezi (John Stuart Mill, mai ales). Nu primeşte o educaţie intelectuală specială în domeniul psihologiei.
Începe cariera în psihologie, lucrând cu Jean Martin Charcot (1825-1893), celebrul psihiatru care a dat maximă strălucire şcolii de la Salpêtrière. Din 1891, Binet activează în laboratorul de psihologie fiziologică de la Sorbona. În anul 1895 fondează cunoscuta revistă *L'année psychologique* (al cărei prim director a fost Henri Piéron), unde publică primele 9 articole de craniometrie, orientare avându-l ca mentor pe Paul Broca.
Binet şi Galton mor în anul 1911. Galton avea 89 de ani şi o îndelungată productivitate intelectuală. Binet avea 54 de ani şi era la apogeul carierei. Un biograf al lui Binet (T.H. Wolf) afirmă în 1973 că Binet a fost tatăl psihologiei experimentale în Franţa şi că el a avut cel mai mare impact în psihologia din SUA, după moartea lui Wundt.

1. Nu căutăm să ignorăm contribuţia celuilalt autor al scalei, Théodore Simon (1873-1961), dar considerăm că cel mai mare merit îi revine lui Binet. Simon a trăit încă cinci decenii după moartea lui Binet, dar nu a mai realizat nici o perfecţionare a testului.

3.1. Bazele teoretice, experimentale şi metodologice ale apariţiei scalei Binet-Simon

Recunoaşterea scalei metrice a inteligenţei, creată de Alfred Binet şi Théodore Simon, ca prototip al testului psihologic se datorează intersectării unor acumulări ştiinţifice apreciabile cu nevoile instituţionale, reflectate în comenzi adresate savantului francez, fapt care a făcut ca testul psihologic să intre în viaţa practică a societăţii. Desigur, nu sunt de neglijat calităţile constructive ale probei, care au servit drept model pentru numeroase teste ulterioare. Alfred Binet, căruia i se atribuie unanim paternitatea metodei testului, a fost o personalitate ştiinţifică de excepţie, apreciere datorată unor realizări teoretice şi practic-experimentale de pionierat. După ce înfiinţează, la Sorbona, primul laborator de psihologie fiziologică, se ocupă intens o perioadă de timp de craniometrie, abandonează aceste preocupări după un perseverent efort de măsurare a unor eşantioane de sute de copii şi, totodată, renunţă la convingerea că există o corelaţie între mărimea craniului şi inteligenţă. Se îndreaptă către mijloacele psihologice de studiu al inteligenţei şi, în anul 1896, reproşează înaintaşilor de a fi acordat o importanţă excesivă proceselor psihice inferioare şi de a fi neglijat aptitudinile superioare (atenţia, imaginaţia, inteligenţa, care diferenţiază mai bine decât simţul tactil sau olfactiv). În 1896, Alfred Binet şi asistentul său de la Sorbona, Victor Henri, publică un articol despre diferenţele individuale, în care argumentează că inteligenţa poate fi măsurată mai curând decât procesele senzoriale elementare, cum ar fi timpul de reacţie, acuitatea senzorială. Deşi are realizări practice indubitabile, savantul francez este considerat în primul rând un teoretician, datorită entuziasmului cu care participă la dezbaterile filosofice ce bulversau domeniul său de referinţă şi datorită interesului teoretic de durată pentru studiul inteligenţei. Astfel, în 1886 publica *Psihologia raţionamentului* (prima sa carte), iar în 1903 *Studiul experimental al inteligenţei*, lucrare în care abjura vechile angajamente intelectuale şi care marchează distanţarea faţă de psihologia de laborator. El concepe experimentarea într-un sens foarte larg, aceasta trebuind să cuprindă un ansamblu variat de instrumente metodologice (chestionar, interviu, anchetă).

Eforturile lui Binet apar în contextul unui puternic curent ştiinţific de natură psihologică, marcat de crearea laboratoarelor psihologice în numeroase ţări, de apariţia revistelor specializate (în Franţa, *Annales médico-psychologiques*, *L'année psychologique*; în Germania, *Philosophische Studien* şi *Zeitschrift für Psychologie*; în SUA, *American Journal of Psychology* şi *Psychological Review*) şi de organizarea congreselor internaţionale de psihologie (primul Congres internaţional de psihologie a avut loc la Paris în anul 1889; al doilea la Londra în 1892; al treilea la München în 1896, la care participă 500 de psihologi din întreaga lume; al patrulea la Paris în 1900). La aceste congrese au

prezentat lucrări J.M. Charcot, Théodule Ribot, Pierre Janet, Théodore Flournoy, Édouard Claparède, William James, Alexandre Bain, Hugo Münsterberg, H. Ebbinghaus. Cu tot avântul preocupărilor psihologice, lucrările teoretice privind elaborarea testului erau aproape inexistente, o excepţie notabilă constituind-o realizările lui Francis Galton, consacrate mai mult mecanismelor psihofiziologice şi timpului de reacţie.

Alfred Binet a fost preocupat permanent de cercetarea psihofiziologică. La începutul carierei, preocupările sale au fost mai largi, axându-se apoi mai ales asupra aspectului cognitiv. În domeniul medical conduce lucrări de hipnotism. Într-un prim studiu, susţine că ar putea să manipuleze senzaţiile şi simptomele subiectului hipnotizat prin mişcarea unui magnet în diverse locuri ale corpului acestuia, realizare ce ar putea avea aplicaţii atât în medicina generală, cât şi, mai ales, în domeniul psihiatriei. Alţi cercetători nu pot reproduce concluziile lui Binet şi atribuie rezultatele sale unui slab control experimental. După discuţii ştiinţifice controversate pe această temă, Binet admite, în final, că rezultatele sale se datorau sugestiei, şi nu puterii magnetului. Acest incident, considerat de Binet un eşec, l-a determinat să demisioneze de la Salpêtrière, în 1890, şi l-a marcat toată viaţa.

Mai târziu îşi îndreaptă atenţia spre evoluţia intelectuală a copilului, având ca subiecţi pentru aplicaţie pe cele două fiice ale sale, în vârstă de 2 ani şi jumătate, respectiv 4 ani şi jumătate, în momentul începerii observaţiilor. Testele create pentru a investiga operaţiile mintale sunt similare celor pe care Jean Piaget le-a creat mai târziu. De exemplu, întreabă care din două grămezi conţine mai multe obiecte şi găseşte că răspunsul era determinat de cantitatea de spaţiu pe care o ocupă obiectele, şi nu de numărul obiectelor aflate în grămadă. Binet se ocupă, de asemenea, de memoria de scurtă durată a fiicelor sale. În cercetările sale foloseşte un număr din probele utilizate de Galton şi Cattell pentru a măsura acuitatea vizuală şi timpul de reacţie. În 1890, publică trei articole descriind cercetările efectuate asupra fiicelor sale şi, ulterior, în 1893, un studiu experimental asupra inteligenţei, valorificând cercetările în cauză.

După ce Binet intră în Laboratorul de psihologie fiziologică de la Sorbona, realizează cercetări referitoare la natura fricii infantile, aria memoriei, fidelitatea mărturiei vizuale, creativitate, similitudinea dintre idei şi exprimarea grafică. În perioada activităţii de la Sorbona, a investigat diferenţele individuale în percepţia petelor de cerneală, înaintea apariţiei faimoasei lucrări a lui Hermann Rorschach. Binet era interesat în mod special de psihologia individuală, de ceea ce era diferit la oameni sub aspect psihofiziologic. În 1896, asistentul său Victor Henri (1872-1940) a scris un articol despre psihologia individuală, unde propune o listă de variabile în care indivizii diferă mai ales în plan intelectual. Ambii au în atenţie o listă de variabile importante şi calea de determinare a gradului în care fiecare variabilă există într-un anumit individ. Cu ajutorul unor variabile izolate şi al unei metode de măsurare disponibile, ei sperau că individualul ar fi posibil de evaluat, într-o perioadă scurtă de timp. Procedurile descrise de Galton şi Cattell erau respinse, deoarece aceştia acordau importanţă proceselor senzoriale, şi nu celor mintale superioare. Concepţiile celor doi cercetători erau respinse şi pentru că acordau o minimă importanţă diferenţelor dintre psihicul infantil şi cel adult. Ca atare, Henri şi Binet îşi propun să studieze abilităţile cognitive în mod direct, şi nu indirect. Ei sunt de acord că variabilele importante ce diferenţiază oamenii sunt procesele superioare complexe, care se transformă o dată cu vârsta.

Variabilele incluse în lista din anul 1896 cuprind memoria, imaginaţia, atenţia, comprehensiunea, sugestibilitatea, plasticitatea, judecata estetică şi cea morală, forţa voinţei şi aprecierea spaţiului vizual.

Aşteptările lui Binet şi Henri de a accede într-un timp scurt la procesele mintale superioare nu s-au împlinit, deoarece administrarea testelor cerea mai multe ore, iar interpretarea rezultatelor necesita, de asemenea, mai multe ore de judecare clinică subiectivă.

Un efect descurajant l-a avut un studiu statistic al testelor lor, efectuat de către Stella Sharp, licenţiată la Cornell University. În 1899, cercetătoarea americană găseşte intercorelaţii foarte scăzute între testele lui Binet şi Henri şi concluzionează (precum Clark Wissler în privinţa testelor lui Cattell) că acestea nu puteau evalua inteligenţa.

Drept urmare, se renunţă la proiectul psihologiei individuale, însă Alfred Binet nu se dă bătut şi foloseşte experienţa câştigată la următorul proiect.

3.2. Evaluarea deficienţei individuale

În 1899, Théodore Simon, intern la o mare instituţie din localitatea Perray-Vaucluse care se ocupa de asistenţa medico-psihologică a copiilor cu retard mintal, îl invită pe Binet să supervizeze cercetarea sa de doctorat. Binet agreează ideea şi vede în aceasta o oportunitate de a avea acces la un mare număr de subiecţi pentru studiu. În acelaşi an, el intră în Societatea liberă pentru studierea psihologică a copilului. Această organizaţie are drept scop obţinerea de informaţii ştiinţifice despre copil, în special despre problemele educaţiei lui. Devine în curând lider al organizaţiei. Ulterior, Binet şi Simon vor fi aleşi într-un grup pe care guvernul francez l-a însărcinat să studieze problema copiilor cu retard mintal din şcolile franceze.

În acea vreme, pentru detectarea retardării mintale erau folosite teste variate create de Galton. Copiii orbi sau surzi erau clasificaţi ca având tulburări mintale.

Unul dintre elementele care au contribuit substanţial la apariţia prototipului testului psihologic a fost, fără îndoială, comanda socială. În anul 1904, când Binet abandonase deja definitiv metodele medicale în favoarea celor psihologice privind studiul inteligenţei, Comission Bourgeoise din Ministerul Instrucţiei i-a comandat un studiu care avea ca obiectiv crearea unui instrument practic de măsurare a inteligenţei, în vederea evidenţierii copiilor retardaţi, cu scopul ca aceştia să urmeze o formă de educaţie specializată, adaptată nivelului lor intelectual. Pentru a onora comanda, Alfred Binet a optat pentru un demers pur pragmatic. Practic, s-a răspuns la o problemă socială de organizare a învăţământului special pentru retardaţi. Comanda socială va sta la baza creării altor teste şi va influenţa, în multe privinţe, evoluţia vieţii ştiinţifice în psihologie. Modul de rezolvare a comenzii sociale şi individuale, în profesiunea de psiholog, precum şi reacţia socială la aplicarea testelor psihologice vor fi la originea multora dintre oscilaţiile în aprecierea valorii testului psihologic şi a metodei testului, în general.

Câteva luni mai târziu, în anul 1905, Binet şi Simon propun o scală metrică a inteligenţei. Cererea socială a fost un declanşator, care l-a determinat pe Binet să pună rapid în formă ideile şi proiectele sale. Din 1890 până în 1905, Binet a căutat semnele exterioare de inteligenţă. Pentru aceasta, el a colaborat cu Simon, care a studiat idioţii şi imbecilii din azilul Perray-Vaucluse. El compară elevii după vârstă şi după eficacitatea şcolară. Aceste comparaţii vizează percepţia lungimilor şi a numerelor, a culorilor, interpretarea desenelor, definiţia obiectelor, memorizarea cuvintelor şi a frazelor, descrierea obiectelor. A elaborat numeroase probe de mici dimensiuni pentru a-şi observa intensiv cele două fiice. Scala metrică a inteligenţei este rezultatul tuturor acestor cercetări. Se poate spune că prima versiune a scalei Binet-Simon s-a bazat pe serioase preparative anterioare de natură teoretică şi metodologică.

Pentru prima scală elaborată, Binet şi colegul său mai tânăr au creat un evantai larg de probe scurte având legătură cu problemele vieţii cotidiene (a urmări cu privirea un obiect în mişcare; a indica diferite obiecte; a numi obiectele dintr-o imagine; a compara linii de lungimi inegale; a repeta trei cifre rostite de examinator; a compara două greutăţi; a repeta o propoziţie de 15 cuvinte; a produce rime; a întoarce ceasul). Descrierea principalelor caracteristici relevă o serie de aspecte constructive şi de situaţii problematice, care vor persista în activitatea realizatorilor de teste. Mai mult de jumătate dintre itemii primei scale erau destinaţi celor cu o accentuată retardare.

Prima versiune a scalei metrice comportă 30 de mici probe, în general reuşite la vârste diferite (nu este dată nici o indicaţie referitor la vârste) şi prezentate în ordinea crescătoare a dificultăţii. Probele alese au legături intrinsece cu funcţiile care, după Binet, definesc inteligenţa. Inteligenţa nu este totuşi definită prin natura sa (eficienţă în funcţiile particulare), ci prin geneza sa: nivelurile sale sunt vârste mintale. Binet, după cum a relevat Zazzo în 1966, nu utilizează expresia „vârstă mintală", ci vorbeşte despre „nivel mintal", însă el poate să fie considerat inventatorul noţiunii de *vârstă mintală*, care caracterizează performanţele unui subiect raportate la performanţele medii ale copiilor de diferite grupe de vârstă. Despre un copil în vârstă de 10 ani (vârstă cronologică) se spune că are 12 ani vârstă mintală, dacă nivelul său de reuşită este echivalent aceluia al copilului mediu de 12 ani. Dintre doi copii având aceeaşi vârstă cronologică, se consideră ca fiind mai inteligent acela care are vârsta mintală mai ridicată. Cu vârsta mintală, Binet utilizează deci referinţa la dezvoltare, pentru a fonda o măsurare de nivel ordinal. Se poate nota că Binet nu raţionează în termeni de creştere mintală (coeficientul intelectual – raportul dintre vârsta mintală şi vârsta cronologică), măsură a vitezei de dezvoltare care va fi introdusă în circuitul ştiinţific de W. Stern, în 1912. Această primă versiune a scalei metrice nu vrea să spună că un instrument permite caracterizarea nivelului de deficienţă şi depistarea debililor. Insuficienţa mintală este definită doar în termeni de întârziere şi de oprire a dezvoltării. Caracterul insuficient al definiţiei, simplă consecinţă a opţiunii făcute pentru a măsura inteligenţa, nu i-a scăpat lui Binet.

În 1904, Binet şi Simon îşi propun doar să creeze teste cu care să diferenţieze copiii normali şi subnormali din punct de vedere intelectual. Deoarece autorii îşi propuneau o metodă psihologică de diagnostic ştiinţific al stărilor inferioare ale inteligenţei, se punea problema unei definiri precise a stării de normalitate şi subnormalitate (*cf.* Roşca, 1972, p. 159). Problema definirii clare a conceptelor utilizate în intenţia diagnostică a testelor psihologice este şi astăzi de actualitate. Binet şi Simon resping opticile mecaniciste în apreciere, atunci când se diferenţiază între oprirea dezvoltării şi deviaţiile de dezvoltare: „Un idiot de 15 ani care este la primele încercări verbale ale unui copil de trei ani nu poate fi complet asimilat unui copil de trei ani, deoarece acesta este un normal, iar idiotul un infirm" (*ibidem*).

Primul pas consta în izolarea unui grup de copii clar diagnosticaţi ca normali şi a altui grup clar diagnosticat ca subnormal.

Al doilea pas consta în testarea ambelor grupe pe diferite căi, în scopul de a distinge un grup de altul pe baza rezultatelor realizate. Binet era convins că cea mai bună cale de a examina diferenţele individuale era în termeni complecşi de procese mintale, prin folosirea unor teste special construite pentru diagnosticarea acestora.

După multe încercări şi erori, Binet şi Simon ajung la primul test de inteligenţă direct, şi nu indirect, prin acuitatea senzorială (aspiraţie declarată de pe vremea când colabora cu Henri).

În anul 1905, cei doi autori realizează o metodă de testare care era mai mult un produs al tradiţiei raţionaliste decât al celei empiriste. Autorii prezintă scala ca fiind o cale validă de distingere a copiilor normali de cei deficienţi, o cale care să înlocuiască criteriile fizice, sociale şi educaţionale folosite la timpul respectiv pentru identificarea copiilor retardaţi, metode considerate mai puţin fidele. În 1905, scala consta din 30 de probe ordonate după gradul de dificultate, pornind de la simpla mişcare a ochilor şi ajungând la definiţii abstracte. Trei dintre aceste probe măsurau dezvoltarea motorie, iar alte 27 erau desemnate să măsoare abilităţile cognitive. Ordonarea după gradul de dificultate a probelor făcea ca pe cele mai dificile să le rezolve copiii mai dotaţi. Scala era administrată copiilor normali şi copiilor cu retard mintal, cu vârste de la 2 la 12 ani. Solicitările scalei se prezintă în felul următor:

1. Demonstrarea coordonării vizuale
2. Demonstrarea apucării (prehensiunii) după contact (apucarea unui cub după atingere)
3. Demonstrarea vederii unui lucru
4. Recunoaşterea hranei (alegerea între lemn şi ciocolată)
5. Căutarea hranei (ca răspuns la ciocolata ambalată în hârtie)
6. Urmărirea comenzilor simple sau repetarea gesturilor
7. Indicarea obiectelor
8. Recunoaşterea obiectelor într-un desen
9. Numirea obiectelor într-un desen
10. Diferenţierea lungimii a două linii
11. Repetarea a trei numere
12. Diferenţierea a două greutăţi
13. Rezistenţa la sugestie
14. Definirea cuvintelor simple
15. Repetarea propoziţiei de 15 cuvinte
16. Stabilirea diferenţei între perechi de obiecte
17. Demonstrarea memoriei vizuale
18. Desenarea unor forme din memorie
19. Demonstrarea memoriei temporale pentru numere
20. Formularea similarităţilor între obiecte
21. Discriminarea rapidă a liniilor
22. Ordonarea a cinci greutăţi
23. Identificarea greutăţii care lipseşte (dintre cele cinci de la testul 22)
24. Furnizarea de cuvinte care rimează
25. Completarea unor propoziţii
26. Construirea unor propoziţii conţinând trei cuvinte date
27. Răspunsul la întrebări (de exemplu *Ce visezi tu ?*)
28. Potrivirea acelor unui ceasornic când acestea au fost schimbate
29. Îndoirea şi tăierea hârtiei
30. Distingerea între termeni abstracţi (precum *trist* şi *plictisit*) (*cf.* Hergenhahn, 1997, pp. 279-280).

Binet şi Simon stabilesc că majoritatea copiilor normali având vârsta de 2 ani şi puţin peste 2 ani pot să răspundă cu uşurinţă la probele numerotate de la 1 la 6. Copiii cu retard slab sau moderat puteau să efectueze unele dintre aceste probe sau chiar pe toate.

Cele mai multe probe, numerotate de la 7 la 15, puteau fi trecute de copiii normali având vârsta între 2 şi 5 ani. Copiii cu retard uşor puteau să treacă fiecare dintre aceste teste; copiii cu retard moderat aveau mari dificultăţi, iar copiii cu retard sever nu au putut să treacă nici una. Probele numerotate de la 16 la 30 puteau fi, în mod obişnuit, trecute de copiii normali având vârsta între 5 şi 12 ani; copiii care prezentau chiar un retard uşor au avut mari dificultăţi cu aceste probe, iar copiii cu retard moderat şi sever nu puteau să treacă nici una.

În prima scală elaborată de Binet şi Simon se poate evidenţia convingerea celui dintâi potrivit căreia inteligenţa nu este o abilitate singulară, ci multiplă, expunând concepţia psihologică a mai multor filosofi raţionalişti. El nu acceptă totuşi nativismul care însoţeşte adesea punctele de vedere raţionaliste şi crede că moştenirea ereditară poate plasa la limita de sus unele abilităţi intelectuale; în acelaşi timp însă, consideră că, de cele mai multe ori, fiecare individ funcţionează la limita de jos a potenţialului său.

O altă convingere a lui Binet este aceea că fiecare poate să-şi mărească potenţialul intelectual şi că acest fapt poate fi un factor de primă importanţă în educaţie.

D.R. Goodenough (1949) a schiţat cele patru motive pentru care scala prezentată în anul 1905 diferă de cele realizate anterior (*cf.* Gregory, 1996, p. 13):

1. Este concepută fără pretenţia de a măsura câteva facultăţi individuale, fiind, mai degrabă, destinată să aprecieze dezvoltarea mintală generală a copilului printr-un grup eterogen de sarcini.
2. Este un test scurt şi practic: presupune mai puţin de o oră pentru administrare şi necesită, de asemenea, puţine echipamente.
3. Testul măsoară direct ceea ce Binet şi Simon au considerat a fi factorul esenţial al inteligenţei, judecata practică. Ei au avut o viziune practică asupra inteligenţei.
4. Itemii au fost ordonaţi la un nivel de dificultate adecvat vârstei.

Este demn de remarcat faptul că, deşi demersul lui Binet a fost unul intelectual, aspectul practic este cel mai interesant la instrumentul de investigaţie pe care l-a lansat. El a considerat că raţionamentul teoretic este inadecvat pentru a realiza o metodă eficientă de departajare a subiecţilor şi, ca atare, a plecat de la convingerea că ştiinţa veritabilă progresează prin inducţie, pornind de la simple experienţe realizate cu scopul de a culege fapte fundamentale, şi nu de a pune la probă teorii complexe. Binet refuză să ofere o interpretare scalei sale de inteligenţă. Spre deosebire de testele precedente concepute să măsoare „facultăţi" ale spiritului, specifice şi independente, scala lui Binet se prezintă sub forma unei colecţii de activităţi diverse. A rămas celebră o frază în care căuta să evidenţieze caracterul empiric al activităţii sale: „Se poate merge până la a spune: puţin importă testele, important este ca ele să fie numeroase" (vezi Gould, 1984, p. 179). Credinţa lui Binet era că, îmbinând un număr suficient de teste de aptitudini diferite, se va ajunge la conturarea într-un rezultat singular al potenţialităţilor generale ale unui copil. Încă din 1905, el încearcă să separe inteligenţa naturală de instrucţie. În realitate, era convins că examinează o stare de fapt mai complexă decât inteligenţa verbal-abstractă, exersată în instruirea şcolară. Creşterea numărului de teste este văzută ca o posibilitate de ieşire din îngustimea conceptuală a acestui tip de inteligenţă, operaţie completată de luarea în considerare a aprecierii datelor oferite de teste în raport cu cele rezultate din analiza biografică a subiecţilor. Acest gen de analiză complexă a datelor rezultate din examenul psihologic presupunea să se acorde o atenţie deosebită formării

prealabile a profesioniştilor din domeniu, aspect în care Binet şi Simon au excelat. Aceeaşi atenţie acordă autorii francezi exigenţelor în privinţa construirii testelor psihologice, care trebuiau să aibă următoarele calităţi:

- să fie apropiate pentru fiecare vârstă de inteligenţă;
- să fie independente de toate condiţiile exterioare şi conjuncturale, în particular de cunoştinţele şcolare;
- să fie egal aplicabile la indivizi de toate naţionalităţile şi de toate culturile;
- să beneficieze de o execuţie simplă şi să nu presupună nici aparate de precizie, nici concentrarea atenţiei pe durată lungă;
- să fie suficient de exact pentru a se putea compara ulterior rezultatele tuturor cercetărilor;
- să fie în stare să dea o măsură globală a inteligenţei (*cf.* Chamboulant, 1951, p. 23).

A doua versiune a scalei metrice apare în 1908 şi diferă esenţial de prima. Dintre cei 58 de itemi, doar 17 erau prezenţi şi în versiunea din 1905, noii itemi creaţi fiind adăugaţi la capătul superior al scalei. Noua versiune permitea diferenţierea şi între copiii normali, ceea ce nu era cazul în cea dintâi. Lista itemilor este clar structurată în funcţie de vârsta medie la care ei sunt reuşiţi, de la 3 la 13 ani. Aceste norme de vârstă permit calcularea vârstei mintale.

Dificultăţile constau în reconstituirea unor propoziţii fragmentate, copierea unui pătrat, execuţia unei secvenţe de trei comenzi. Câteva dintre situaţiile prezentate erau absurdităţi pe care copiii trebuiau să le detecteze şi să le explice. Un astfel de item avea următorul conţinut: „Corpul unei nefericite fetiţe a fost găsit tăiat în 18 bucăţi. Se crede că a fost vorba despre o sinucidere" (vezi Gregory, 1996, p. 13).

Această forrmulare căreia trebuia să-i fie detectat absurdul amuza copiii francezi, dar i-a supărat pe cei americani, evidenţiindu-se în acest fel importanţa factorului cultural în aprecierea inteligenţei.

Inovaţia majoră a constat în introducerea conceptului de *nivel mintal*. Scopul scalei revizuite nu mai era acela al unei simple distingeri între copiii normali şi copiii prezentând o anumită retardare, ci realizarea unei separări între nivelurile inteligenţei la copiii normali de 3 până la 13 ani. Dacă 75% sau mai mulţi dintre copiii de o anumită vârstă rezolvau un test particular, acest test era sigur pentru un anumit nivel de vârstă. De exemplu, cei mai mulţi copii de 4 ani puteau să copieze un pătrat, dar nu un caro. Autorii au stabilit că o mică parte dintre copiii de 3 ani puteau să copieze un pătrat, mare parte dintre cei de 4 ani (75% sau mai mult) puteau să realizeze această copie şi aproape toţi copiii de 5 ani puteau s-o facă. Puteau să determine câţi dintre copii sunt sub medie şi câţi deasupra, la vârsta de 5 ani. Cei de 5 ani care realizau corect o dificultate a testului rezervat celor de 5 ani erau consideraţi normali. Dar, în cazul în care copiii treceau doar testele tipice pentru 4 ani, aceştia erau consideraţi ca având o inteligenţă sub medie. Dacă treceau şi testele rezervate celor de 6 ani, erau consideraţi a avea o inteligenţă peste medie. Cu alte cuvinte, nivelul de inteligenţă al unui copil era determinat de valoarea normei pe care copilul o realiza.

Pentru standardizarea testului au fost examinaţi 300 de copii normali, între 3 şi 13 ani. Pentru vârsta bazală (3 ani), folosirea unui singur test a fost considerată insuficientă. Pentru fiecare 5 teste trecute deasupra nivelului bazal se adăuga în diagnosticare 1 an în privinţa nivelului mintal.

Solicitările scalei revizuite din anul 1908 erau următoarele :

3 ani A arăta nasul, ochii, gura ; a enumera părţile componente ale unui desen ; a repeta două cifre ; a repeta o construcţie verbală de şase silabe ; a-şi spune numele de familie.

4 ani A indica sexul său ; a numi cheia, cuţitul, banul ; a repeta trei cifre ; a compara două linii.

5 ani A compara două cutii de greutăţi diferite ; a copia un pătrat ; a repeta o construcţie de zece silabe ; a număra patru monede simple ; a reface un joc din două piese.

6 ani A repeta o construcţie de 16 silabe ; a compara două figuri din punct de vedere estetic ; a defini, prin folosire individuală, obiecte familiare ; a executa trei sarcini simultan ; a-şi spune propria vârstă ; a distinge între dimineaţă şi seară.

7 ani A indica lacunele într-o figură ; a număra cele zece degete ; a copia un triunghi şi un romb ; a repeta cinci cifre ; a descrie un desen ; a număra 13 monede simple ; a numi patru monede.

8 ani A citi un text oarecare, reţinând două lucruri în memorie ; a număra nouă monede (trei simple, trei duble) ; a numi patru culori ; a număra descrescător de la 20 la 0 ; a compara două obiecte din memorie ; a scrie după dictare.

9 ani A indica data completă (ziua, luna şi anul) ; a enumera zilele săptămânii ; a utiliza definiţii superioare ; a memora şase elemente, după lectura unui text oarecare ; a ordona cinci greutăţi ; a împărţi 20 de monede în grupuri de câte patru.

10 ani A enumera lunile ; a numi nouă monede ; a localiza trei cuvinte în două fraze ; a răspunde la trei întrebări care solicită inteligenţa ; a răspunde la cinci întrebări care solicită memoria.

11 ani A critica fraze ce conţin absurdităţi ; a localiza trei cuvinte într-o frază ; a spune mai mult de 60 de cuvinte în trei minute ; a da definiţii abstracte ; a pune cuvintele în ordine.

12 ani A repeta şapte cifre ; a găsi trei rime ; a repeta o construcţie verbală de 26 de silabe ; a interpreta desene.

13 ani A face un decupaj ; a completa un triunghi (*apud* Huteau şi Lautrey, 1995, pp. 22-23).

În 1911, se definitivează şi se publică cea de-a treia revizie. Ea va diferi puţin de versiunea din 1908 (lărgirea scalei spre nivelurile superioare, simplificarea calculului nivelului mintal). Pentru fiecare nivel de vârstă, scala conţine cinci teste. Noua scală reformată cuprindea şi date normative pentru adulţi (15 ani). Se introduce o nouă metodă de calculare a scorurilor, care acorda o cincime de an pentru fiecare subtest trecut dincolo de nivelul bazal. Se insistă asupra faptului ca nivelul mintal să nu fie luat ca o măsură absolută a inteligenţei, ci doar ca un punct de reper servind nevoilor de ordonare a subiecţilor.

Această variantă a scalei permitea o măsurare mai rafinată a inteligenţei. De exemplu, dacă un copil de 8 ani rezolva dificultăţile testului corespunzând vârstei sale, el putea fi considerat normal. Uneori însă, copilul dorea să efectueze şi testele specifice pentru vârsta de 9 ani. Noua procedură permitea să capete un punctaj suplimentar faţă de cel al vârstei sale.

Nivelul intelectual al copilului putea fi exprimat în termeni de vârstă intelectuală sau vârstă a inteligenţei. Această vârstă corespundea celor mai mari dificultăţi pe care copilul a reuşit să le rezolve. Binet a subliniat extrema prudenţă în interpretarea vârstei intelectuale a copilului. În primul rând, el a observat că era absolut normal ca un copil să aibă o vârstă intelectuală cu doar un an mai puţin decât vârsta cronologică, acesta având probabil mici tulburări la şcoală. Copiii la care vârsta intelectuală era cu doi ani sau mai mult în urma vârstei lor cronologice întâmpinau dificultăţi în programul şcolii obişnuite şi aveau nevoie de o atenţie specială. Însă, chiar şi atunci când se obţin performanţe slabe la teste, nu se impune, în general, considerarea şcolarilor respectivi ca fiind deficienţi mintal. Înainte să se aplice o etichetă, administratorul testului trebuia să se asigure că şcolarul era sănătos şi motivat atunci când el a efectuat testul şi că avea cunoştinţe despre cultura franceză pentru a înţelege implicaţiile culturale din test.

Coeficientul de inteligenţă. În 1911, renumitul psiholog german William Stern, emigrat ulterior în Statele Unite, introduce noţiunea de *coeficient de inteligenţă*. Pentru Stern, vârsta mintală a unui copil era determinată de performanţa sa pe scala Binet-Simon. Stern a sugerat crearea unui *coeficient de inteligenţă*, rezultat din raportul dintre vârsta mintală şi vârsta cronologică. Denumirea s-a propagat în rândul psihologilor şi nu numai al lor. Astfel, dacă un copil de 7 ani trece toate probele tipice acestei vârste, coeficientul său de inteligenţă este $\frac{7}{7}$, deci 1,00. Dacă un copil de 7 ani trece doar testele tipice pentru un copil de 5 ani, coeficientul de inteligenţă este $\frac{5}{7}$, deci 0,71. În 1916, Lewin Madison Terman sugerează ca acest coeficient să fie înmulţit cu 100, pentru a elimina exprimarea zecimală. De asemenea, Terman a abreviat coeficientul de inteligenţă IQ, formulă foarte răspândită astăzi. Combinarea sugestiilor lui Terman şi Stern conduce la formula :

$$IQ = \frac{VM}{VC} \times 100,$$

unde VM = vârsta mintală, VC = vârsta cronologică.

Binet s-a opus folosirii coeficientului de inteligenţă, considerând că inteligenţa este prea complexă pentru a putea fi reprezentată printr-un simplu termen sau număr. Simplificarea lui Stern s-a impus, în ciuda opoziţiei lui Binet.

Se poate aprecia că cei doi savanţi francezi au dezvoltat un instrument de măsurare a inteligenţei simplu, uşor de administrat. La începutul primului război mondial a fost cel mai folosit test din lume.

3.3. Concepţia lui Binet asupra menirii scalei de inteligenţă

În primul rând, Binet vedea în scala sa un procedeu pentru identificarea copiilor care au nevoie de educaţie specială. Credinţa puternică a lui Binet era că şcolarii care au obţinut scoruri slabe puteau să-şi mărească randamentul în mod considerabil, dacă li se acorda o atenţie specială. Deşi Binet credea că ereditatea poate să impună o limită potenţialului intelectual, el considera că un mare număr de oameni pot să ajungă la un potenţial intelectual deosebit , dacă există o stimulare corespunzătoare. Pe Binet îl irita faptul că o bună parte dintre profesorii din acea vreme aveau convingerea çă performanţele intelectuale erau înnăscute. În felul acesta, celor care erau angrenaţi în procesul de învăţământ sub formă de beneficiari li se putea induce ideea că aveau o inteligenţă scăzută. El nu accepta viziunea nativismului extrem al inteligenţei. De aici şi ideea *ortopediei mintale* care poate pregăti pentru şcoală copiii dezavantajaţi. Ortopedia mintală constă în exerciţii menite să amelioreze voinţa, atenţia şi să disciplineze toate abilităţile necesare pentru educaţia efectivă în clasă. Binet credea că, prin ortopedie mintală, copiii vor fi învăţaţi „cum să înveţe".

Alfred Binet s-a interesat de problema măsurării inteligenţei, dar preocuparea sa principală a fost crearea unei „psihologii individuale", adică a unei „psihologii a individualităţii", în termenii folosiţi în prezent de psihologia personalităţii sau de psihologia diferenţială (ultimul termen a fost propus de W. Stern în 1900). Metoda testelor a fost prezentată ca una dintre metodele acestei „psihologii individuale". Pentru Binet, studiul funcţionării psihologice generale şi studiul variabilităţii interindividuale sunt strâns legate, a doua fiind un mijloc pentru prima. Alte metode ale psihologiei individuale sunt *metoda cazurilor anormale* (căutarea incidenţei unei slăbiri sau unei hiperfuncţionări a unui proces asupra altor procese) şi *metoda experimentală* (constând în a face să varieze un proces referitor la diferenţele individuale şi a observa consecinţele asupra altor procese). El abordează problema individualităţii prin cogniţie. Obiectivul său este de a căuta în ideaţie ceea ce este personal în fiecare dintre noi. În opera sa se găsesc două definiţii ale inteligenţei. Una s-a emis în termeni de adaptare socială: inteligenţa este „simţul comun, simţul practic, iniţiativa, facultatea de adaptare" (1905 - *apud* Huteau şi Lautrey, 1995, p. 20). Altă definiţie a fost emisă în termeni de funcţionare psihologică: inteligenţa constituie rezultanta acelor mari funcţii distincte care sunt comprehensiunea, invenţia, direcţia şi cenzura (1909). Una dintre problemele centrale ale psihologiei individuale este determinarea variabilităţii şi a modului de organizare a acestor funcţii.

Binet a fost foarte critic în privinţa primelor instrumente de investigaţie psihologică denumite *teste*. Analizând primele realizări americane şi germane, el publică, împreună cu V. Henri, un studiu intitulat *La Psychologie individuelle* şi reproşează primelor probe psihologice cu titlul de teste că sunt nepractice şi incomplete. Într-adevăr, aceste realizări sunt, în practică, foarte puţin comode: sunt lungi, plicticoase şi presupun un material complex, fiind mai ales incomplete, în măsura în care se interesează cu prioritate de procesele elementare. Binet nu dispreţuieşte procesele elementare (a consacrat o serie de publicaţii studiului sensibilităţii tactile - esteziometriei), recunoscând că acestea pot fi măsurate cu o precizie mult mai mare decât procesele superioare. El observă că diferenţele de precizie sunt, în parte, iluzorii, deoarece variabilitatea proceselor superioare este

mult mai puternică decât aceea a proceselor elementare. Insistă asupra interesului de studiu al acestor procese: ele permit definirea „trăsăturilor caracteristice" ale individului. Dar Binet nu doar critică, ci şi propune propriul său program. Pentru a fi utilizabil, testul nu trebuie să depăşească o oră şi jumătate de solicitare a subiectului. El trebuie să fie constituit din exerciţii variate, pentru a stimula atenţia, şi nu trebuie să facă apel la instalaţii speciale. Pentru a fi complet, testul trebuie să solicite procesele mintale superioare. Binet şi Henri furnizează o listă a respectivelor procese: 1) memoria (la care enumeră diverse aspecte); 2) natura imaginilor mintale; 3) imaginaţia; 4) atenţia; 5) facultatea de a înţelege; 6) sugestibilitatea; 7) sentimentele estetice; 8) sentimentele morale; 9) forţa musculară (considerată forţă a voinţei); 10) abilităţile motrice şi privirea.

Binet nu este singurul care a propus un test destinat diagnosticării proceselor psihice superioare. Ebbinghaus, ale cărui lucrări clasice asupra memoriei au fost publicate în 1885, elaborează în 1897 un test de completare a frazelor, după ce J.A. Gilbert, în Statele Unite, elaborase un test de memorie şi de sugestibilitate. Hugo Münsterberg include în bateria sa un test de memorie şi de vocabular (1891). Galton a construit probe de asociere verbală şi de completare de fraze (găsirea unui cuvânt omis într-o frază, plecând de la context).

Realizările lui Binet le-au eclipsat pe acelea ale precursorilor. Cel puţin patru inovaţii permit înţelegerea succesului de care s-a bucurat testul lui Binet: în primul rând, caracterul sistematic al demersului (ansamblul testului este cel care vizează procesele superioare sau doar unele dintre părţile sale); în al doilea rând, pertinenţa şi bogăţia situaţiilor de observaţie alese, rezultând din 15 ani de cercetări asupra dezvoltării cognitive; în al treilea rând, trebuie menţionată şi inovaţia care constă în utilizarea diferenţelor de vârstă (vârsta mintală) pentru a descrie diferenţele individuale relative la o vârstă cronologică constantă. În sfârşit, în construcţia lor, Binet şi Simon au reuşit să fie eficienţi în plan practic, testul permiţând depistarea apreciată ca satisfăcătoare a debililor şi o prognosticare a reuşitei şcolare. Din anul 1908, succesul scalei Binet-Simon este fulgerător. O primă adaptare s-a realizat în SUA, de către Henry Goddard, în anul 1909. O a doua, cea mai răspândită, este datorată lui Lewis Terman, fiind cunoscută sub numele de „Terman-Stanford", apărută în 1916. Ea permite calculul unui IQ de tip Stern şi dezvoltă măsura inteligenţei la adult. În fine, o a treia adaptare este prezentată în anul 1937 sub numele de „Terman-Merrill". Ultima adaptare americană a testului francez va fi preluată în Franţa, unde va trebui să vină anul 1949 pentru ca testul lui Binet să fie reetalonat sau anul 1966, pentru realizarea unei noi versiuni, intitulată *La Nouvelle échelle métrique de l'intelligence* (bine cunoscuta *NEMI* a lui René Zazzo). O dată cu realizarea lui Binet, crearea testelor psihologice este relansată pe noi baze. Testele create ulterior se inspiră mult, la început, din scala Binet-Simon, diversificându-se mai apoi. Bazele teoretice şi metodologia construcţiei lor vor evolua considerabil. Se vor manifesta şi anumite derive de la principiile de construcţie.

Cu tot acest standard înalt de exigenţe (care este apreciat parţial şi astăzi), aplicaţiile practice au început să scoată în evidenţă o serie de deficienţe, dintre care amintim următoarele:

- testele prezintă inconvenientele şi dificultăţile inerente probelor psihologice individuale;
- ele avantajează copiii dotaţi în plan verbal (calitate mult mai des întâlnită la cei care provin din mediile sociale favorizate);

- testele nonverbale de la primele vârste trebuiau revizuite, deoarece punerea în practică le-a indicat inadecvarea (cazul aplicării de către Décrolly în Belgia);
- testele în care intervenea noţiunea de *număr* s-au dovedit inadecvate şi incorect gradate;
- pentru copiii străini sau cu lacune în privinţa limbajului, s-a arătat că probele de limbaj trebuie să intervină mai puţin sau chiar deloc.

În istoria psihologiei şi în istoria ştiinţei, în general, figura lui Alfred Binet este cu totul remarcabilă, prin ţinuta deontologică impecabilă. El a fost permanent preocupat de acurateţea activităţii sale şi de etica cercetării ştiinţifice, de prevenirea posibilităţii de a deforma rezultatele unor măsurători prin propria sugestibilitate. Vorbeşte despre „uimitoarea maleabilitate a datelor cantitative, obiective când ele sunt în serviciul unei idei preconcepute" (*apud* Roşca, 1972, p. 177). De asemenea, aminteşte cu o rară dezinvoltură de mărirea inconştientă, cu bună-credinţă, a preciziei măsurătorilor volumului cefalic al celor consideraţi inteligenţi şi micşorarea în cazul neinteligenţilor (în cadrul cercetărilor de craniometrie efectuate în colonia Vaucluse, azil psihiatric în care Simon era intern).

Binet era şi deosebit de scrupulos în ceea ce priveşte utilizarea cifrelor pentru a exprima realităţi complexe. De aceea, refuza să desprindă semnificaţia numărului pe care el îl atribuia fiecărui copil. „Inteligenţa este prea complexă pentru ca un singur număr să poată să o definească" (*apud ibidem*, p. 181). Numărul este, în concepţia lui Binet, un ghid empiric grosier, conceput într-un scop practic bine delimitat. El consideră că scopul scalei concepute de el şi Simon, mai ales pentru gradele inferioare ale inteligenţei, este de a detecta, în scopul de a acorda ajutor şi a ameliora, nu de a impune limite.

Pentru cei care îşi propuneau să aplice cu precădere teste de inteligenţă, Alfred Binet a insistat asupra următoarelor principii esenţiale, care, din păcate, au fost uitate în unele cazuri:

- Rezultatul obţinut este un simplu instrument, şi nu exponentul unei teorii a intelectului uman. El nu defineşte şi nu precizează nimic înnăscut sau permanent. Savantul francez nu agrea denumirea de inteligenţă pentru ceea ce măsura testul pe care l-a creat, cu toate că proba a fost consacrată cu acest nume.
- Scala nu este decât un ghid empiric grosier, util pentru a semnala educatorilor copiilor uşoare arierări sau tulburări dislexice, având nevoie de un ajutor specializat. Scala nu era un instrument pentru a clasa copii normali.

Indiferent de cauza dificultăţii pe care o întâlnesc copiii astfel selecţionaţi, acestora trebuie să li se acorde ajutor printr-un învăţământ specializat. Proba nu poate fi un pretext pentru a considera aceşti copii inapţi congenital, pe baza rezultatelor slabe.

Am menţionat anterior că evoluţia testului psihologic este strâns legată de evoluţia curentelor psihologice în ansamblu, de progresul evident al celorlalte metode psihologice de investigaţie şi de evoluţia vieţii sociale, care a avut momente şi fenomene favorizante şi defavorizante. În perioada în care s-a creat şi perfecţionat scala Binet-Simon, asistăm la dezvoltarea altor teste psihologice, proces evoluând în interferenţă cu creaţia savanţilor francezi. Creatorii de teste, contemporani lor, au fost influenţaţi, mai ales după finalizarea scalei lor de inteligenţă.

3.4. Continuarea principiilor lui Alfred Binet

Ideile lui Binet şi Simon în materie de construcţie a testului psihologic s-au continuat în trei direcţii:

- extinderea noţiunii de *vârstă a inteligenţei* către cea de *vârstă a dezvoltării* (realizările având ca autori pe A. Gesell, în 1925, Ch. Buhler şi H. Hetzer, în 1932, L. Schrank-Danziger, în 1953);
- „alunecarea" (glisarea) noţiunii globale a inteligenţei în mai multe funcţii independente una de alta, de exemplu testele de inteligenţă analitică ale lui L.L. Thurstone (1938), D. Wechsler (1939, 1949), R. Meilli (1930) şi R. Amthauer (1953). Modelul lui Wechsler a fost continuat prin prelucrările germane ale Institutului de Psihologie din München, şi anume HAWIE (1955) şi HAWIK (1956);
- crearea testelor nonverbale de inteligenţă, de exemplu jocuri de montaj şi compunere, *„form board"* şi „puzzle" (după modelele create de către E. Seguin, încă din 1866), lansate de R. Pintner şi D.G. Paterson în 1917, forme de labirint (după S.D. Porteus, 1924), compoziţii de cuburi (după S.C. Kohs, 1923), manipulări de abstracţii şi figuri geometrice (matricea progresivă standard, creată în 1938 de J.C. Raven).

Dacă principiile lui Binet ar fi fost urmate, s-ar fi evitat abuzuri ştiinţifice majore. Acestea apar ca efect, înainte de toate, a două iluzii, şi anume reificarea şi ereditarismul. Eroarea reificării constă în considerarea notelor IQ ca reprezentând o realitate unică, numită „inteligenţă", măsurabilă şi localizabilă în craniu. Eroarea ereditarismului constă în afirmarea faptului că ereditatea creează cadre imuabile de performanţă individuală, precum şi în confuzia între ereditar şi inevitabil, între ereditatea din cadrul unui grup şi ereditatea între grupuri. În acest ultim caz, eroarea constă în convingerea că, dacă ereditatea explică un anumit procentaj de variaţie între indivizi în interiorul unui grup, ea trebuie să explice un procentaj similar de diferenţă a mediilor IQ între grupuri (albi şi negri, negri şi galbeni etc.).

4. Instituţionalizarea psihologiei şi extensia câmpului de aplicare a testelor

Se poate afirma, fără riscul de a greşi, că psihologia se instituţionalizează o dată cu trecerea de la secolul al XIX-lea la secolul XX, în SUA mai repede decât oriunde în lume, şi că testele psihologice joacă un rol esenţial în acest proces. Înainte de a fi obiecte de aplicaţie, testele psihologice au fost mai întâi obiecte de cercetare.

Se constată o preocupare, în rândul universităţilor din cele mai dezvoltate state ale lumii, pentru învăţământul psihologic, în general, şi pentru testul psihologic, în particular.

Responsabilitatea primului departament de psihologie, independent de filosofie şi de disciplinele umaniste, este încredinţată la University of Pennsylvania lui James McKeen Cattell, promotorul testelor psihologice în Statele Unite. Începând din anul 1892, există interes pentru teste în majoritatea celor 22 de universităţi americane de la acea vreme, o consecinţă imediată a acestui fapt fiind crearea laboratoarelor de psihologie. Laboratoarele

nu aveau ca singură menire construirea, aplicarea şi verificarea testelor psihologice. În 1910, American Psychological Association fusese deja condusă de 13 preşedinţi, dintre care 11 sunt cunoscuţi pentru cercetările lor asupra testelor mintale. Atât în SUA, cât şi în Europa de Vest, în anii premergători primului război mondial, testele rămân un apanaj universitar, dar, ulterior, se vor extinde la diverse domenii de aplicaţii.

Primul este domeniul educativ, în care se manifestă nevoia de diagnostic al capacităţilor individuale. În 1906, Henry Herbert Goddard este primul psiholog care obţine un post cu normă întreagă în afara mediului universitar. El va conduce departamentul de cercetare al unei instituţii de asistenţă a debililor mintal. În 1912, P. Bouvet şi É. Claparède fondează la Geneva Institutul „Jean-Jacques Rousseau". Primul psiholog şcolar, bine cunoscutul Cyril Burt, este acceptat de London Country Council, în anul 1913.

Aproximativ în aceeaşi perioadă, apar aplicaţii practice ale psihologiei (în particular, ale testelor) în domeniile muncii şi sănătăţii. Începând cu 1890, Francis Galton, Alfred Binet şi Hugo Münsterberg (în general, considerat fondator al psihologiei industriale) au investigat, prin intermediul testelor, persoane care aspirau să ocupe anumite funcţii. Câţiva ani mai târziu, vor fi efectiv întreprinse aplicaţii sistematice. În vederea construirii unor teste adecvate selecţiei profesionale, unul dintre cei mai cunoscuţi constructori de teste, J.M. Lahy, studiază activitatea dactilografelor, în 1905, a conducătorilor de tramvai, în 1908, şi, ulterior, pe cea a linotipiştilor. El propune, în anul 1910, un test de atenţie difuză destinat să selecţioneze conducătorii de tramvai, test folosit şi astăzi de către laboratoarele de psihologie din domeniul transporturilor. Münsterberg se preocupă, de asemenea, de selecţia conducătorilor de tramvai, în 1910, de aceea a standardistelor, în 1914, şi a tehnicienilor din telecomunicaţii, în 1915. În 1909, întreprinde o vastă anchetă vizând evidenţierea caracteristicilor psihologice corespunzând la diferite posturi de muncă, reluând un proiect al său mai vechi, din anul 1891.

Pe lângă selecţie, începe să se dezvolte şi orientarea profesională. În 1908, sunt deschise cabinete de orientare profesională în Belgia, de către A.G. Christiaens şi O. Decroly, în SUA de către F. Parsons (care va publica în 1909 prima carte despre orientarea profesională). Un cabinet de acelaşi tip este deschis la Paris, de către Chaintreau, în 1910. În Elveţia, Claparède şi Bouvet se preocupă, şi ei, de acest domeniu. Orientarea profesională în diferite meserii va deveni un câmp privilegiat pentru aplicarea testelor. Pentru fondatorii primelor cabinete de orientare profesională, principala problemă de rezolvat era aceea a diagnosticului aptitudinilor cerute de diferite meserii.

În Statele Unite ale Americii, în anul 1909, este angajat primul psiholog folosit într-un cabinet de psihiatrie. Acestuia i se vor încredinţa, în esenţă, sarcini de diagnostic prin intermediul testelor, iar respectivul serviciu specializat va rămâne permanent apanajul psihologilor din acest sector.

După cum se observă, deşi sunt notabile, primele aplicaţii ale testelor rămân totuşi foarte limitate.

4.1. Contribuţii americane la dezvoltarea testului psihologic

Dezvoltarea testului psihologic în SUA poartă amprenta stilului de viaţă american şi idealului american, rezumabile în câteva tendinţe clar reliefate:

- tendinţa de a situa performanţa în prim-plan; de aici, marea anvergură a aplicării testelor, din dorinţa de a depăşi tot ce se făcuse până atunci. Putem exemplifica afirmaţia anterioară prin gigantismul unor realizări fără precedent şi foarte greu de depăşit. Ne referim la milioanele de subiecţi examinaţi cu un set de probe, la crearea unor instrumente psihologice, care au necesitat reevaluări foarte costisitoare în timp şi bani. În acea vreme, creatorii americani de teste psihologice aveau fizionomia spirituală a unor apostoli, care au făcut un crez greu de zdruncinat din necesitatea răspândirii unei noi speranţe intelectuale, şi anume aceea că potenţialul psihic uman poate fi evaluat cu precizie;
- tendinţa de a face profitabilă din punct de vedere economic orice fel de îndeletnicire. James McKeen Cattell obiectivează această tendinţă, încă de la revenirea sa în Statele Unite, prin crearea unei organizaţii comerciale denumite Psychological Corporation, care a pus la dispoziţia domeniului educativ şi industrial serviciile oferite de reprezentanţii săi. Dar Jastrow, în 1892, şi Wissler, în 1901 (ambii colaboratori ai lui Cattell), cunosc gustul amar al eşecului (*cf.* Reuchlin, 1967, pp. 56-57). Imediat după experienţa cu *Army Alpha* şi *Army Beta*, din timpul primului război mondial, apar o puzderie de firme care caută să exploateze comercial aceste noi instrumente de investigare psihologică. Respectiva optică economică a adus avantaje, dar şi mari necazuri pentru psihologi;
- tendinţa pragmatică, vizând folosirea cu ingeniozitate în construcţia testului psihologic a oricărui element care ar putea face mai eficient şi valoros noul produs ştiinţific. Optica constructivă era următoarea: de ce să se elaboreze un instrument nou, când poate fi perfecţionat unul valoros, deja existent? Primii constructori americani au perfecţionat şi adaptat realizarea lui Binet şi Simon.

Pionierii construcţiei testului psihologic în America sunt H.H. Goddard, L.R. Terman, R.M. Yerkes şi Leta Stetter Holingsworth.

4.1.1. *Problema „moronilor"*

Psihologului Henry Herbert Goddard i se datorează „importarea" şi traducerea în America (în anul 1908) a scalei Binet-Simon şi reificarea IQ-ului, considerat inteligenţă înnăscută. O bună parte din sursele bibliografice îl prezintă ca pe un excentric care s-a preocupat de clasificarea retardaţilor mintal. Dacă în privinţa idioţilor şi imbecililor funcţionează un consens general în stabilirea sau recunoaşterea diagnosticului, în privinţa *„micilor" deficienţi mintal*, care puteau să primească o formaţie şi să fie utili în societate (şi care stabileau legătura dintre normal şi patologic), exista şi continua să existe un pericol pentru edificiul taxonomic din domeniul psihologic. Cei cu vârsta mintală între 8 şi 12 ani, denumiţi *débiles* de către francezi şi *feebleminded* de către americani şi englezi, constituiau subiectul unor dispute interminabile. Exista o mare tentaţie la taxonomişti de a confunda crearea unui nume cu rezolvarea unei probleme. Henry Herbert Goddard a fost unul dintre cei care au avut această tentaţie.

Henry Herbert Goddard (1866-1957) s-a născut în New England, într-o familie de quakeri. Obţine *bachelor's degree* şi *master's degree* la Haverford College. După ce a fost profesor de colegiu şi director timp de 7 ani, s-a înscris pentru un stagiu de doctorat în psihologie la Clark University, pentru a da curs interesului său în educaţie şi psihologie. Goddard îşi susţine disertaţia de doctorat (care investiga factorii psihologici implicaţi în vindecarea prin credinţă) sub supervizarea lui G. Stanley Hall.

După acordarea titlului de doctor, în 1899, Goddard este primul acceptat ca profesor în domeniul psihologiei la Pennsylvania's West Chester State Teacher's College şi, în 1906, devine director de cercetare la Training School for the Feebleminded, în Vineland, New Jersey. Traduce toată opera lui Binet şi Simon în engleză şi, după moartea celui dintâi, în 1911, devine susţinător al concepţiilor lui Binet în măsurarea inteligenţei. Deşi iniţial era sceptic faţă de scala Binet, el o găseşte foarte eficientă în clasificarea copiilor în termeni de grad de retardare. Cu toate că foloseşte procedurile de testare ale lui Binet, Goddard acceptă mai curând viziunea Galton-Spearman--Cattell despre natura inteligenţei decât pe cea a lui Binet. Aplică scala Binet-Simon, tradusă de el, la Vineland School şi la 2.000 dintre cursanţii şcolilor publice din New Jersey. Este şocat de constatarea că mulţi dintre cursanţii şcolii publice aveau performanţe intelectuale sub norma vârstei lor. Credinţa sa în caracterul ereditar al inteligenţei derivă din faptul că, în urma observării copiilor din Vineland, găseşte că, adesea, fraţii şi surorile acestora erau *feebleminded* (termen predilect folosit de Goddard pentru a-i denumi pe cei cu intelect limitat).

Acest energic director de cercetare la Vineland Training School for Feebleminded Girls and Boys din New Jersey a inventat un termen, şi anume cel de *moron*, semnificând „debil uşor" (după cuvântul grec care înseamnă „prost", „stupid"), termen care s-a impus în dezbaterile ştiinţifice ale vremii.

Goddard era de acord cu Binet, socotind testul un instrument de selecţionare a persoanelor care se aflau sub nivelul normalului, dar, spre deosebire de acesta, care refuza să privească notele ca desemnând inteligenţa, considera că ele desemnează o entitate unică. El îşi propunea să identifice indivizii aflaţi la limita normalului (pentru a le determina limitele potenţialului intelectual) şi să-i izoleze (pentru a le „scurta" educaţia), iar toate acestea în scopul de a evita continuarea degradării unui filon american puternic (sever ameninţat din exterior – prin imigraţie –, respectiv din interior – de către cei „slabi de spirit").

În 1911, examinează 1.547 de copii normali şi ajunge la concluzia că cei consideraţi *feebleminded* (cu patru ani vârstă mintală mai puţin decât vârsta cronologică) erau în proporţie de 3%, considerată alarmantă. Atitudinea sa ştiinţifică are la bază o clasificare uniliniară a deficienţei mintale, mergând de la idioţi la imbecili şi apoi la debili, precum şi ipoteza potrivit căreia evoluţia urmează un drum liniar, care poate fi măsurat de o scală unică, mergând de la ceea ce este primitiv la ceea ce este avansat şi care dă cel mai bine seama de variaţiile observate. Întrebarea este dacă multitudinea de cauze şi fenomene care se regrupează sub numele de deficienţă mintală pot fi ordonate într-un mod rezonabil pe o scală unică. Aceasta ar presupune că fiecare persoană, clasată prin utilizarea doar a scalei respective, posedă o anumită cantitate dintr-o substanţă unică numită inteligenţă.

Iată câteva fenomene care pot să ducă la obţinerea unei cantităţi limitate din această „substanţă unică": slaba dezvoltare mintală generală, dificultăţi specifice de învăţare datorate traumatismelor neurologice locale, dezavantaje datorate mediului, diferenţelor culturale, ostilităţii examinatorilor şi, nu în ultimul rând, un instrument inadecvat de investigare a inteligenţei. Cauze posibile pot fi: modurile ereditare de funcţionare, patologia genetică accidentală netransmisibilă, leziunile creierului datorate unei boli a mamei în timpul sarcinii, malnutriţia fetusului şi a sugarului, diferite deficienţe ale mediului, care intervin în prima copilărie şi ulterior. În ciuda acestor cauze, Goddard

numea „moroni" toate persoanele care aveau vârsta mintală între 8 şi 12 ani. La început, singura soluţie întrevăzută de Henry Goddard era plasarea lor în instituţii speciale, unde să fie riguros controlate, să fie făcute „fericite", prevăzând nevoile lor limitate, şi, mai ales, să fie împiedicate să se reproducă.

H.H. Goddard socotea că principalul agent al conduitei umane îl reprezintă un proces mintal unitar pe care noi îl numim inteligenţă, proces condiţionat printr-un mecanism nervos înnăscut. Esenţialul conduitelor umane indezirabile este atribuit inteligenţei. Legătura dintre imoralitate şi inteligenţă era una dintre temele predilecte ale eugeniştilor din acea perioadă. Extensia în domeniul ideologiei sociale era următoarea: „Nu poate să existe egalitate socială, atâta timp cât există o accentuată inegalitate mintală".

Goddard a fost contemporan cu primele manifestări entuziaste prilejuite de redescoperirea lucrărilor lui Mendel şi a primelor fundamente ale eredităţii. În acest stadiu al dezvoltării geneticii, toate trăsăturile umane se comportau precum greutatea, înălţimea, culoarea etc. Deci existenţa organelor celor mai complexe ale corpului uman se datora acţiunii unei singure gene, iar variaţiile anatomice şi cele ale comportamentului rezultau din forma dominantă sau recesivă a genei respective. Actualmente, se ştie, practic, că toate caracteristicile mai importante ale corpului nostru sunt elaborate prin interacţiunea a numeroase gene, între ele şi cu mediul exterior. Goddard a încercat să reconstituie arborele genealogic al asistaţilor din şcoala pe care o conducea la Vineland şi a tras concluzia că debilitatea se supune regulilor mendeliene ale eredităţii. În 1914, era de părere că inteligenţa normală apare ca o caracteristică unitară, transmisă în manieră pur mendeliană.

Ulterior, Goddard va considera că reproducerea moronilor trebuie doar încetinită, precizând că ar fi suficientă colonizarea, în locul mutilării prin sterilizare. O altă cale de a limita ponderea moronilor era aceea de a limita drastic numărul celor care veneau din străinătate, deci de a reduce sever imigrarea.

Testele mintale şi imigrarea. Între 1901 şi 1913, milioane de indivizi au emigrat din Europa în SUA, iar o parte dintre aceştia erau suspectaţi de tulburări mintale. Se punea problema elaborării unor proceduri de detectare a lor. Mişcarea eugenistă din acea vreme era destul de puternică şi avea adepţi în multe state americane. În 1912, comisia de imigrare îl invită pe Goddard la Ellis Island, ca să observe şi să examineze imigranţii. Goddard susţinea că poate să spună care dintre imigranţi avea deficienţe mintale, pe baza observării caracteristicilor fizice, însă, pentru a fi sigur, apelează şi la scala Binet-Simon. Pe baza rezultatelor la teste, unii imigranţi erau etichetaţi deficienţi mintal şi erau deportaţi. Încurajat de experienţa de aici, Goddard primeşte subvenţii pentru a-şi completa studiul şi efectuează, cu ajutorul a două asistente ale sale, examinări pentru a repera debilii din rândul imigranţilor. Calitatea de examinator a femeilor era considerată net superioară bărbaţilor, datorită intuiţiei lor înnăscute. Goddard consideră că experienţa de durată în domeniul examinării psihologice duce la achiziţionarea unui simţ special, permiţând o mare capacitate de diferenţiere. El a mers atât de departe încât a specificat ţările europene pentru care procentajul de imigranţi cu deficienţe mintale era cel mai înalt. Cele două colaboratoare ale lui Goddard au examinat 35 de evrei, 22 de unguri, 50 de italieni şi 45 de ruşi, după ce instrumentul cu care lucrau a fost tradus în limbile respective. Eşantionul se stabilea la întâmplare, deci aleatoriu, cu toate că funcţionarii de la biroul „Imigrări" îi eliminaseră deja pe cei consideraţi deficienţi. Pentru a reechilibra această situaţie, Goddard şi colaboratoarele sale i-au lăsat deoparte pe cei consideraţi „evident normali", rămânând de examinat marea masă a imigranţilor de nivel mediu. Astăzi, operarea cu această masă de prejudecăţi uimeşte, dar atunci procedura era considerată

obiectivă. Imigrantul mediu era apreciat, din start, sub standard. Aplicarea testelor lui Binet a dus la rezultate surprinzătoare: 83% din evrei, 80% din unguri, 79% din italieni şi 87% din ruşi erau uşor debili, adică aveau o vârstă mintală inferioară celei de 12 ani pe scala Binet-Simon. Desigur, aceste rezultate sunt de-a dreptul stupefiante: cum poate cineva să susţină că trei sferturi dintre reprezentanţii unei naţiuni sunt debili, şi mai ales pe baza examinării unor imigranţi care, prin motivaţia lor, constituie, evident, un eşantion nereprezentativ? Ipoteza că testele n-au fost bine înţelese de participanţi era respinsă, deoarece, de exemplu, evreilor le-a dat explicaţii cineva care vorbea idiş.

În general, Goddard conchidea că între 40 şi 50% dintre imigranţi erau moroni. Explicaţiile sunt următoarele: traducerea originală, pe care Goddard a făcut-o scalei Binet-Simon, nota foarte sever examinaţii şi clasa în categorii de debili subiecţi consideraţi, în mod obişnuit, ca fiind normali. Când Terman emite versiunea Stanford-Binet, constată că versiunea lui Goddard atribuia note net mai mici decât făcea instrumentul său. Din 104 adulţi normali după scala lui Terman, 50% erau consideraţi debili dacă erau examinaţi cu scala folosită de Goddard. Ca şi în activitatea sa anterioară, Goddard îşi asuma faptul că performanţele la teste ale imigranţilor se datorau în special inteligenţei moştenite, şi nu educaţiei culturale sau experienţei personale (toţi aceşti factori au fost găsiţi ulterior ca influenţând performanţele la test).

Însă imigranţii au efectuat testul în circumstanţe speciale, de cele mai multe ori testele fiind administrate prin intermediul translatorilor. În urma eforturilor lui Goddard, rata deportărilor creşte cu 350% în 1913 şi cu 570% în 1914, faţă de media anilor anteriori. Forţa de muncă ieftină era pierdută pentru SUA. Goddard era mulţumit. În anii următori, îşi va schimba radical credinţele, îmbrăţişând câteva dintre ideile lui Binet.

Cauzele acestor severe concluzii erau mai mult decât evidente: avem de-a face cu grupuri de oameni puternic stresaţi, majoritatea necunoscând limba engleză, oameni care îndurase oboseala unei lungi călătorii la clasa a treia, care erau în majoritate săraci, care nu fuseseră niciodată la şcoală şi, în mod firesc, n-au ţinut niciodată un stilou sau măcar un creion în mână.

Examinarea se desfăşura în felul următor: la puţin timp după ce coborau în serii din vapor, una dintre colaboratoarele „intuitive" ale lui Goddard alegea câte un imigrant, îl ruga să se aşeze, îi întindea un creion şi îi cerea să reproducă pe hârtie o figură pe care i-a arătat-o câteva secunde şi pe care apoi i-o sustrage privirii. Eşecul, în acest caz, era mai degrabă explicat de condiţiile în care se desfăşura testarea, de slăbiciune şi frică, de confuzie, decât de prostia înnăscută. Procedurile de examinare l-au determinat pe David Hothersal să-i acorde lui Goddard un loc de cinste într-un capitol consacrat abuzurilor în folosirea testelor de inteligenţă (Hothersal, 1983).

Pe parcurs, Goddard îşi „îndulceşte" poziţia în funcţie de complicaţiile practice care puteau apărea datorită măsurilor restrictive luate împotriva moronilor. Totuşi, continua să susţină necesitatea ca guvernul să ia măsuri pentru asigurarea efectuării procesului de testare psihologică în porturi.

Studiul asupra familiei „Kallikak". Goddard decide să investigheze relaţiile dintre fundamentul familial şi inteligenţă cu mai multă grijă, fapt pentru care întreprinde lungi studii genealogice. În 1911, a administrat scala Binet-Simon lui Deborah Kallikak, care a stat la Vineland School din 1897. Numele „Kallikak" fusese inventat de Goddard, alăturând cuvintele greceşti *kalos* (bun) şi *kakos* (rău). Deşi vârsta cronologică a lui Deborah era de 22 de ani, performanţa rezultată în termeni de vârstă mintală era de

9 ani, rezultând un IQ de 41. Goddard născoceşte termenul *moron* pentru a denumi nivelul intelectual a lui Deborah. El identifică strămoşii acesteia până la revoluţia americană, moment în care Martin Kallilak senior avusese o relaţie sexuală cu o bufetieră *feebleminded* din care a rezultat Martin Kallikak jr. După terminarea armatei, vârstnicul Martin s-a căsătorit cu o fată respectabilă, cu care a avut şapte copii. Tânărul Martin se căsătoreşte şi el şi are zece copii. În analiza lui Goddard, descendenţii bătrânului Martin şi ai „fetei respectabile" reprezintă partea bună a moştenirii lui Deborah, iar descendenţii tânărului Martin, partea rea.

Goddard găseşte că fiii vârstnicului Martin nu au fost nici unul *feebleminded*, pe când cinci dintre copiii tânărului Martin prezentau deficienţe mintale. Pe vremea lui Goddard, o convingere foarte răspândită era că indivizii consideraţi *feebleminded* erau cauza comportamentelor imorale, criminale şi antisociale, iar Goddard arăta că majoritatea descendenţilor tânărului Martin erau hoţi de cai, prostituate, alcoolici, condamnaţi, părinţi de copii ilegitimi sau cu tulburări de ordin sexual. Din sutele de descendenţi ai bătrânului Martin doar trei au avut deficienţe mintale, iar unul a fost considerat „liber sexual". Printre descendenţii bătrânului Martin erau doctori, avocaţi, educatori şi alţi indivizi prestigioşi.

Prezentăm genealogia familiei Kallikak (*apud* Mărgineanu, 1973, p. 218):

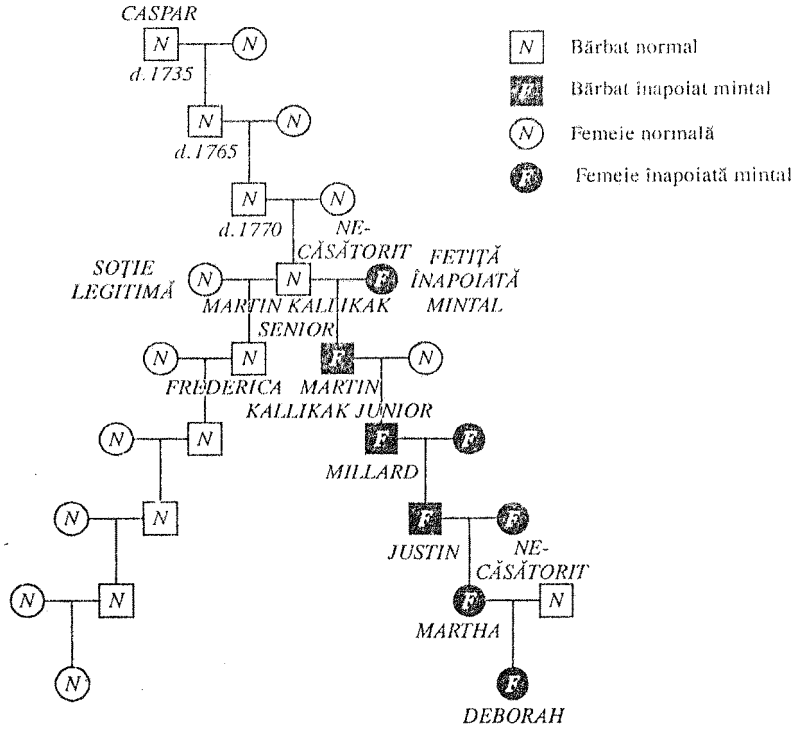


Figura 1. *Genealogia familiei Kallikak*

Goddard a raportat descoperirile sale în *The Kallikak Family, a Study in the Heredity of Feeble-Mindedness*, în anul 1912. Cercetările sale au folosit ca suport pentru credința galtoniană, potrivit căreia inteligența era determinată genetic. În frunte cu Goddard, câțiva lideri științifici solicitau ca locuitorii americani care prezentau tulburări mintale să fie sterilizați și separați de restul societății.

Nu mai puțin de 20 de state adoptă legi de sterilizare și, ca urmare, unii indezirabili au fost sterilizați. Într-o serie de state americane, legile de sterilizare au fost în vigoare până în anii '70. Asemenea decizii i-ar fi făcut plăcere lui Goddard.

Studiile realizate de H.H. Goddard acordau mai multă încredere intuiției decât unui examen riguros, iar fotografiile care însoțesc cartea dedicată familiei „Kallikak" s-au dovedit a fi trucate prin îngroșarea liniilor negre ale trăsăturilor, înfățișându-i pe acești hibrizi cu un aspect sumbru și rigid, sugerând înapoierea mintală. Specialiștii considerau cazul foarte interesant din punctul de vedere al manipulării fotografice. Nu este singurul caz de falsificare a datelor de către unii autori „furați" de prejudecățile lor.

În 1928, găsim o schimbare frapantă în principiile lui Goddard:

- *moronii*, adică persoanele cu debilitate ușoară, nu sunt incurabili;
- debilii ușori nu necesită internarea într-o instituție.

La sfârșitul carierei, Goddard deja credea că a trecut la inamic. În final, el agreează ideea că tratamentul potrivit pentru indivizii care aveau scoruri mici la testele de inteligență era educația specială, nu segregarea sau sterilizarea (însă avea deja multe victime la activ).

4.1.2. *Lansarea coeficientului de inteligență înnăscut*

L.M. Terman este una dintre personalitățile marcante ale psihologiei americane, nu numai prin contribuția la evoluția teoriei și construcției testului psihologic, ci și prin contribuția la viața științifică din domeniul psihologiei, în Statele Unite. A fost președinte al American Psychological Association.

Lewis Madison Terman (1877-1956) s-a născut la 15 ianuarie 1877, fiind al doisprezecelea din cei 14 copii ai unei familii de fermieri din centrul statului Indiana. Merge la o școală cu o singură clasă și își completează cele 8 clase la vârsta de 12 ani. La vârsta de 9 ani, un negustor de cărți frenologice făcuse o analiză frenologică familiei Terman și lui îi prezisese un viitor de excepție, fapt ce l-a stimulat să aspire la o viață diferită de cea de la fermă. Într-adevăr, Terman și-a urmat vocația precoce, devenind o personalitate de prim rang în istoria metodologiei psihologice. La vârsta de 15 ani, părăsește ferma și se înscrie la Central Normal College în Dunville, Indiana. La 17 ani, începe să predea într-o școală rurală. După șapte ani de la părăsirea fermei, obține trei confirmări *undergraduate*: în artă, în științe și în pedagogie. Următorii trei ani au fost foarte activi pentru Terman. Devine director de colegiu, soț și tată. În 1901, se înscrie la Indiana University unde urmează masteratul în pedagogie, după care aspiră la un post de profesor. Primește oferta pentru studii de doctorat la Clark University. Cu sprijinul financiar al familiei, acceptă oferta și, în curând, începe studiile cu G. Stanley Hall, care conducea un seminar, în cadrul căruia doi studenți prezentau opera sa unui număr de 30 de studenți psihologi, pedagogi, filosofi și educatori. Prezentarea studenților și discuțiile care urmau reprezentau ceva ce Terman considera a fi un bun rezumat.

Terman nu și-a elaborat lucrarea de doctorat sub îndrumarea lui Hall, deoarece era interesat în testarea psihologică, iar Hall era prea puțin entuziasmat de această temă. Sub supervizarea lui Edmund C. Sanford, Terman a izolat un grup de studenți *bright* (strălucitori) și un grup de studenți *dull* (cu intelect de limită, prostănaci) și a încercat să determine ce tipuri de teste pot fi folosite la diferențierea membrilor celor două grupuri. Disertația lui Terman se intitula *Genius and Stupidity. A Study of the Intellectual Processes of Seven „Bright" and Seven „Stupid" Boys*. Terman avea să spună mai târziu că tot interesul carierei sale a căpătat formă în perioada anilor petrecuți la Clark University.

După ce obține doctoratul la această universitate, în 1905, Terman se îmbolnăvește de tuberculoză și ajunge la concluzia că cel mai potrivit este să aleagă un climat cald, în care să poată lucra. Din acest motiv, acceptă un post la Colegiul principal din San Bernardino, California. Un an mai târziu, acceptă un post de profesor în domeniul studierii copilului și în pedagogie la Los Angeles State Normal School (mai târziu se angajează la University of California, din Los Angeles). În 1910, Terman ocupă un post la departamentul de educație de la Stanford University, unde mai apoi își desfășoară întreaga carieră. Primește Catedra de la departamentul de psihologie în anul 1922, unde rămâne până la pensionarea sa, în 1942.

Este o coincidență faptul că, o dată cu sosirea lui Terman la Stanford, apare în SUA scala de inteligență a lui Binet și Simon (în traducerea lui Goddard). Terman începe imediat să lucreze pe această scală și își dă seama că nu poate să o utilizeze cu acuratețe, fără modificări, pe copiii din SUA.

Într-adevăr, Terman și-a urmat vocația precoce, devenind o personalitate de marcă în istoria metodologiei psihologice. Henry Goddard a fost, fără îndoială, primul psiholog care a introdus scala Binet în Statele Unite, dar Terman rămâne principalul artizan al generalizării folosirii sale.

Testele Stanford-Binet. Atunci când administrează scala Binet-Simon copiilor din SUA, Terman realizează că rezultatele erau neuniforme. Scorul mediu al copiilor la diverse vârste era mai scăzut decât vârsta cronologică a grupului testat. De exemplu, observă că itemii de la scala Binet-Simon erau prea ușori pentru vârsta de 5 ani și prea dificili pentru vârsta de 12 ani. Această situație făcea ca media vârstei de 5 ani să fie artificial mai înaltă, iar cea de 12 ani artificial mai scăzută. Într-un studiu efectuat împreună cu studentul *graduate* H.H. Childs, Lewis Terman înlătură o serie din itemii existenți în scala Binet-Simon și adaugă noi itemi până când media scorului copiilor devine 100, nu mai dificil decât vârsta lor. Această medie a fiecărei vârste de grup testat, media vârstei mintale, era egală cu media randamentelor vârstei cronologice. Terman și Childs publică prima lor revizie a scalei Binet-Simon în 1912, iar în 1916 Terman publică singur a doua revizie. Versiunea din 1916 va primi numele de *Stanford-Binet*.

În 1916, Terman adoptă noțiunea de *coeficient de inteligență* a lui Stern și sugerează că raportul poate fi înmulțit cu 100 pentru a îndepărta zecimalele, acesta numindu-se *IQ*. Lansarea scalei, care l-a făcut pe Terman bogat și celebru, a fost urmată de o serie de revizuiri. Autorul, profesor universitar la Stanford University, a dat acestei revizuiri numele

de Stanford-Binet, nume care a consacrat-o până astăzi. Ulterior, efortul său se conjugă cu cel al lui Maud Merrill, de asemenea profesor universitar, fapt care face ca scala să fie reținută și sub numele de *scala Terman-Merrill.*

Poziția lui Terman în legătură cu caracterul ereditar al inteligenței. Terman credea că inteligența era, în mare măsură, moștenită. Asemenea lui Goddard, era convins că inteligența scăzută era cauza celor mai multe crime și a altor forme de comportament antisocial. În 1916, considera că o persoană proastă nu poate fi o persoană morală. Terman era impresionat de Binet și a împrumutat mult de la acesta, dar viziunea sa asupra inteligenței era mult mai asemănătoare cu cea a lui Galton. Terman era atât de impresionat de Galton încât a publicat un portret intelectual al acestuia, în care estima că IQ-ul lui Galton era de aproape 200 (apreciere făcută în 1917). Afirmația lui Terman potrivit căreia IQ-ul este o măsură validă a inteligenței native nu a rămas neschimbată. Printre cei mai aspri critici ai săi se numără jurnalistul Walter Lippmann, cu care Terman a dezbătut problema ineității inteligenței într-o serie de articole apărute în *The New Republic*, în 1922-1923. Terman a validat Stanford-Binet, corelând testele de performanță cu estimările profesorilor asupra inteligenței și rangului școlar. El găsește corect corelații înalte în fiecare studiu realizat, însă acestea nu erau surprinzătoare, întrucât trăsăturile și abilitățile pe care școala și profesorii le apreciau ca fiind înalte la științe erau aceleași trăsături și abilități care produceau scoruri înalte la scala Binet-Simon. Prezența corelațiilor respective nu însemna nicidecum că performanțele academice pot fi prezise din performanțele la test.

Dacă într-adevăr testul măsura inteligența nativă, Terman nu a determinat-o niciodată.

În literatura românească de specialitate, scala a fost prezentată amănunțit de către Mariana Roșca (1972, pp. 168-186). Practic, scala Stanford-Binet este o suită de revizuiri ale scalei Binet-Simon, care a „încremenit" în 1911, o dată cu moartea lui Alfred Binet. Până în prezent se cunosc patru revizii ale scalei inițiale, realizate în anii 1916, 1937, 1960 și 1986[1]. Scalele au fost, decenii la rând, metoda principală de examinare individuală a inteligenței, fapt care a făcut ca ea să fie preferată chiar de către psihodiagnosticienii francezi. Încercăm să sistematizăm principalele planuri de evoluție în cadrul revizuirilor succesive.

1) Planul deontologiei. Acest plan este onorat prin respectul față de efortul înaintașilor, față de meritul științific al lui Binet și Simon. Practic, cei doi profesori americani, Galton și Terman, au creat un nou test după modelul francez, dar au păstrat în denumirea probei numele acestuia. Dintre cele 54 de teste care alcătuiesc scala Binet-Simon, după revizia din 1911, au fost păstrate doar 19 (32 fiind restructurate și 3 înlăturate) și au fost adăugate circa 40 de probe noi, dintre care 27 originale (restul fiind preluate de la Keithmann și Healy-Fernald). Un alt domeniu, cu totul și cu totul de excepție, în care s-au remarcat psihologii americani, a fost cel al grijii pentru adecvarea probelor psihologice și al respectului pentru complexitatea acestui segment al creativității umane, care este construcția testului psihologic. Numai revizia din 1916 a cuprins în examinare mii de copii normali, sute de copii deficienți și supradotați, precum și sute de adulți. Pentru a anula diferențele dintre rezultate datorate examinatorilor, aceștia au beneficiat de o formare specială, operațiune care a durat mai mult de o jumătate de an.

1. Asupra acestei ultime forme nu avem documentație.

2) Planul concepției caracteristicilor psihice măsurate. Versiunea din 1916 cuprindea explicații privind folosirea standardelor de vârstă, natura funcțiilor intelectuale și conceptul măsurării inteligenței generale, care funcționa ca adaptabilitate mintală la probleme noi. Pentru asigurarea funcționalității acestei noi concepții erau folosite grupe paralele de teste. Componentele vechi ale testelor, care au fost găsite ca având o valoare discriminatorie inadecvată, au fost înlăturate.

Aprecierea calității măsurate depășea stadiul funcțiilor elementare în favoarea unui întreg combinatoriu, conceput ca sumă a proceselor ce sunt implicate în adaptarea mintală. Se furnizează o lungă listă de atribute operative ale inteligenței generale pe care le relevă testul Stanford-Binet: comprehensiunea limbajului, memoria, întinderea vocabularului, orientarea în spațiu și în timp, coordonarea vizual-motorie, judecata, asemănările și deosebirile, raționamentul aritmetic, cunoașterea obiectelor familiare, îndemânarea și ingeniozitatea în situații practice dificile, capacitatea de a distinge absurdităților, rapiditatea și bogăția asociațiilor de idei, abilitatea de a reconstitui un ansamblu de forme (puzzle) sau idei plecând de la elemente disociate, facultatea de generalizare, aptitudinea de a descoperi o lege pornind de la fapte legate între ele. Faptul că modul de calculare a vârstei mintale presupune acordarea aceluiași număr de luni (ale vârstei mintale) pentru diferite probe are la bază concepția potrivit căreia diferitele componente ale complexului numit inteligență au o pondere relativ egală. În 1937, Lewis Terman și Maud Merrill publică *Measuring Intelligence*, o formă extinsă și amănunțită de revizie a scalei Stanford-Binet. Cea de-a doua revizie se bazează pe o mare acumulare de informații asupra comportamentului la diferite tipuri de teste mintale și a interrelaționării abilităților testate. Anumite tipuri de teste prezintă corelații mai înalte, cu estimări independente ale trăsăturilor. Se introduce noțiunea de *vârstă standard* a performanței, cu aserțiunea că inteligența generală este o trăsătură care se dezvoltă o dată cu vârsta. Se făcea precizarea expresă că, prin intermediul scalei, se măsura o abilitate generală, și nu una specifică sau de grup. Forma scalei din 1937 solicită o gamă largă de abilități, prezentând componente picturale și manipulative cu referire la analogii, opoziții, comprehensiune, similarități și diferențe, completări verbale și picturale, absurdități, memorie pentru material cu sens și pentru cifre.

În cercetările lui Terman erau însă prezente și o serie de prejudecăți naționale și rasiale[1]. El subscria la ideea conform căreia etnologia arată că progresul rasial este strâns legat de capacitatea de tratare a conceptelor și a raporturilor matematice. Era un partizan al calităților înnăscute și al preeminenței acestora asupra educației. Terman devine avocatul utilizării universale a testului psihologic, „dovedind" prin cifre avantajele sociale ale izolării retardaților și ale plasării oamenilor acceptabili biologic în funcții convenind nivelului lor mintal. Consideră, mai ales la începutul activității, că examinatorii săi vor determina pragul minim de inteligență necesară pentru a reuși în fiecare branșă de activitate. Practic, el se vede apostolul unei noi ordini sociale. Terman „interzicea" prin teoriile sale accesul la profesiunile de prestigiu și la funcțiile bine remunerate al persoanelor având un IQ inferior lui 100 și susținea că orice „reușită substanțială" cere probabil un IQ superior valorii de 115 sau 120. Susținea că populația cu performanțe mai mici de 75 pe scala IQ este sortită unor activități necalificate, pe când cea cu un IQ

1. Informații amănunțite în Stephen Jay Gould, *La mal-mesure de l'homme*, Éditions Ramsay, Paris, 1984, pp. 211-234.

mergând de la 75 la 85, unor activităţi calificate. Celor cu un IQ între 75 şi 85 le erau recomandate formaţii şi stagii profesionale adaptate. Fără aceste formări, persoanele în chestiune vor avea tendinţa de a părăsi şcoala şi de a se încadra în asociaţii cu un pronunţat caracter bolşevic. Cercetând IQ-ul în diferite profesii, el afirmă că se operase deja o repartiţie imperfectă în societate.

Terman afirmă că, deşi societatea are nevoie de o mare masă de oameni pentru a-şi porni maşinile, sănătatea acesteia depinde de rolul proeminent al câtorva oameni rari, de geniu, care au un IQ ridicat. Împreună cu colaboratorii săi, a publicat o serie de cinci volume intitulată *Genetic Studies of Genius*, în care încearcă să definească şi să înţeleagă persoanele situate la vârful performanţei prin aplicarea scalei Stanford-Binet. El întreprinde o încercare retrospectivă de a măsura IQ-ul principalilor artizani ai istoriei umane, oameni de stat, militari şi intelectuali. James McKeen Cattell publicase deja o listă de clasare a mii de actori ai istoriei. O colaboratoare a lui Terman, Catherine M. Cox, reduce această listă la 282 de personaje şi adună toate informaţiile biografice importante din tinereţea lor pentru a le putea estima IQ-ul. Un IQ numit A1 evolua de la naştere până la vârsta de 16 de ani, iar alt IQ, numit A2, mergea de la 17 până la 26 de ani. Colaboratoarea lui Terman solicită unui număr de cinci experţi, dintre care făcea parte şi Terman, studierea dosarelor şi evaluarea celor două IQ-uri. Trei dintre cei cinci experţi cad la înţelegere să acorde IQ-ului A1 un punctaj în jur de 135 şi IQ-ului A2 un punctaj apropiat de 145. Ceilalţi acordă note divergente, unuia valori net superioare, iar celuilalt net inferioare. Metoda utilizată în studiu era complet inadecvată, fără nici un corectiv posibil. Dosarele întocmite de Catherine Cox au fost constituite din liste cu fapte ieşite din comun din copilăria personalităţilor studiate. Se pleacă de la cifra de bază 100, care se măreşte pe baza fiecărui element notabil al dosarului. Un IQ slab traduce o absenţă a datelor, iar unul ridicat – o listă substanţială de atribute. Totuşi, Catherine Cox stabileşte şapte niveluri de fiabilitate, al şaptelea fiind „simplă ipoteză", nefondându-se pe nici un fel de date. Cantităţile de informaţii sunt foarte diferenţiate şi procedurile nu sunt de multe ori decât expresia prejudecăţilor autoarei.

Munca empirică a lui Terman măsura ceea ce statisticienii numesc „varianţa internă a grupului" (*within-group variance*), adică diferenţele de cotaţie în cadrul unei populaţii unice (toţi copiii unei şcoli, de exemplu). El a arătat că rezultatele elevilor la testele de la vârsta lor tânără rămân constante, în general, în clasarea lor în raport cu alţi copii, de-a lungul anilor. Terman atribuia cea mai mare parte a acestor diferenţe inegalităţii „darurilor naturale", afirmând (fără dovezi) preeminenţa naturii asupra achiziţiilor ulterioare. Aceasta semnifică, în fapt, că între grupe există diferenţe înnăscute, aserţiune mult mai contestabilă. Şirul de erori este apoi agravat prin confuzia între geneza adevăratei patologii şi sursele variaţiilor comportamentului normal. Este, de asemenea, o eroare să se treacă de la diferenţele din cadrul unui grup la diferenţele dintre grupuri. Terman adopta teza ereditaristă asupra raselor şi claselor sociale şi face din confirmarea lor scopul principal al operei sale.

Lewis Terman avea totuşi cunoştinţă de fragilitatea argumentelor sale în favoarea ineităţii inteligenţei. Revizia probei Stanford-Binet, din 1937, este atât de diferită, în concepţie, de scrierile de început încât cu greu îşi poate imagina cineva că au fost scrise de acelaşi autor. De această dată, cuvântul „ereditate" nu mai este deloc utilizat. Termenii folosiţi pentru a explica diferenţele sunt cei referitori la ambianţă. Vechilor curbe publicate despre IQ-ul diferitelor clase sociale li se aduc corecturi explicative, în care se afirmă că diferenţele sunt prea mici pentru a putea furniza informaţii de viitor asupra

vieţii indivizilor. Contribuţia fiecărei cauze nu este clar precizată. Terman nu emite o opinie definitivă asupra diferenţelor dintre copiii din mediul urban şi cei din mediul rural, dar afirmă că singura ipoteză pe care trebuie să o supună verificării este de ordin ambiental. Ca şi în cazul lui Goddard, observăm la Terman o „îndulcire" a concepţiilor anterioare referitoare la ereditarism.

3) Planul comportamentelor care pun în evidenţă caracteristicile măsurate. Poate cea mai delicată parte din elaborarea unui test o reprezintă găsirea unei serii de comportamente care să înglobeze caracteristica sau componente ale caracteristicii exprimate generic. Operaţionalizarea conceptuală este o activitate care ţine mai mult de partea artistică şi intuitivă a profesiei de psiholog. Astfel, la proba Stanford-Binet, nivelul limbajului (care ocupă aproximativ 25% din ansamblul probei) presupune:

- evidenţierea vocabularului activ şi pasiv;
- asocierea continuă a cuvintelor;
- găsirea unor cuvinte care să rimeze cu cele existente;
- rearanjarea cuvintelor într-o propoziţie;
- construirea unei propoziţii folosind cuvintele date;
- conversia verbală a unei imagini.

Nivelul de dezvoltare a gândirii (29% din ansamblul probei) este pus în evidenţă prin:

- stabilirea asemănărilor şi diferenţelor dintre două sau mai multe obiecte;
- reconcilierea noţiunilor opuse;
- capacitatea de înţelegere a unei situaţii cotidiene;
- rezolvarea unei probleme concrete;
- operarea cu numere (de la o anumită vârstă).

Starea memoriei (21% din total) este evidenţiată prin:

- reproducerea imediată a unei serii formate din două până la nouă numere (începând de la 9 ani, în sens invers);
- reproducerea unor desene, după îndepărtarea lor;
- memorarea de imagini;
- memorarea comenzilor şi a comisioanelor;
- reproducerea unei istorioare.

Probele cu caracter manipulativ (5% din totalul scalei) sunt concretizate în:

- construcţii din cuburi;
- înşiruirea de mărgele;
- executarea unui nod;
- parcurgerea unui labirint.

Starea discriminării perceptive (restul până la 100%) este evidenţiată prin:

- găsirea lacunelor;
- găsirea figurilor geometrice corespunzătoare modelului;
- compararea lungimii unor beţişoare.

4) Planul construcţiei şi selecţiei componentelor testului. Planul propriu-zis al construcţiei a fost cel mai bulversat de intervenţiile succesive. Revizuirile repetate ale scalei sunt o dovadă a conştientizării dinamicii variabilelor ce influenţează rezultatul unui test psihologic. Activitatea de revizie a scalei din 1916 a presupus un efort intens timp de 10 ani, perioadă în care au fost verificate sute de probe ale formelor „M" şi „L"

de mai târziu. În general, tehnica de construire a unui test trebuie revăzută ori de câte ori apare o inadecvare în privinţa capacităţii discriminative, a capacităţilor psihometrice sau atunci când teoria care stă la baza construcţiei sale a suferit modificări de profunzime. La revizia din 1937, au fost revăzute anumite tehnici de construcţie, iar anumite proceduri de construcţie au fost reevaluate, dată fiind importanţa lor pentru înţelegerea caracteristicii esenţiale a scalei.

Instrumentul era elaborat pentru măsurarea *inteligenţei generale*. El era caracterizat printr-o mare varietate de componente eterogene, unele dintre ele contribuind mai mult la total decât altele, însă fiecare contribuind suficient de mult la total pentru a justifica includerea sa în scală. IQ-ului i-a fost stabilită media de 100 pentru fiecare nivel succesiv de vârstă. Pentru a o stabili, a fost necesar ca populaţia să fie selectată pentru fiecare nivel de vârstă şi să se aloce componentele nivelului potrivit de vârstă. S-a pus problema selectării componentelor în aşa fel încât acestea să poată fi mostre reprezentative corecte de comportamente inteligente. Pentru aceasta, selecţia respondenţilor urma să furnizeze standarde şi norme pentru fiecare comportament. Căutarea de material potrivit pentru teste a produs sute de componente, selectate pentru meritul lor în furnizarea situaţiilor standard în care comportamentul inteligent era observat. Aşadar, componentele erau selectate pentru corelaţia lor înaltă cu un criteriu acceptat al inteligenţei.

S-a obţinut pentru fiecare componentă o proporţie critică între vârsta mintală a subiecţilor care au rezolvat un anumit item şi a acelora care nu l-au rezolvat.

În 1937, după selecţia a 209 forme provizorii de probe „L" şi 199 de forme provizorii de probe „M", au fost reţinute 258 de probe, câte 129 pentru fiecare formă, după ce analiza a fost efectuată pe 3.184 de subiecţi. Au fost analizaţi factorii statistici implicaţi în determinarea contribuţiei probelor la scorul final. Cele care contribuiau în mare măsură la obiectivitatea în punctare şi varietatea de conţinut erau reţinute.

Solicitările au fost diversificate şi adaptate, iar îmbinarea dintre partea verbală şi cea nonverbală a constituit o permanentă preocupare. Factorilor extratest, care influenţau semnificativ o serie de randamente, li s-a găsit o explicaţie. Laborioasa activitate de revizuire a lui Terman şi Merrill este o întărire a ideii potrivit căreia unei probe psihologice, asemenea unei plante gingaşe, trebuie să i se ofere permanent îngrijiri atente. Acumularea de date trebuie să fie o activitate continuă şi sistematică. În selecţia testelor, s-a ţinut seama de mai multe criterii, cele mai importante fiind validitatea, fidelitatea, obiectivitatea cotării, economia de timp în aplicare, corectare şi interpretare, interesul stârnit subiecţilor, precum şi asigurarea varietăţii probelor. Pentru a putea selecta un test în scală, era nevoie, de asemenea, ca diferenţa dintre vârsta cronologică medie a celor care reuşesc şi vârsta cronologică medie a celor ce nu reuşesc la respectivul test să fie semnificativă.

Deşi sunt de o mare varietate, testele unei scale au fost în aşa fel selecţionate încât să existe o corelaţie satisfăcătoare cu ansamblul scalei (fapt ce întăreşte ideea examinării unei aptitudini generale, dominante în raport cu factorii specifici existenţi). Această exigenţă de selecţie antrenează a serie de contradicţii, pe care le vom expune pe parcursul prezentei lucrări.

Revizuiri ale probei s-au făcut şi după stingerea din viaţă a autorului principal. Revizia din anul 1960 nu are aceeaşi anvergură ca aceea din anul 1937. Din cele două forme „L" şi „M" s-a alcătuit o singură scală L-M, pe baza testelor care s-au dovedit viabile. În ultima variantă, probele verbale ating un nivel prea scăzut, iar cele de memorare mecanică – unul prea ridicat.

5) Planul standardizării, normării şi stabilirii criteriilor de apreciere a testului psihologic sau planul psihometric. Planul strict psihometric, cel referitor la calitatea instrumentului de măsură, este unul privilegiat. S-a acordat atenţie obiectivităţii, fidelităţii şi validităţii. Probele s-au repetat la mici intervale de timp (forma test-retest a fidelităţii), au fost corelate în special cu rezultatele activităţii şcolare, s-au întocmit şi s-au analizat ierarhii ale valorii corelaţiilor cu diverse valori de raportare. Eşantionării şi perturbării provocate de aceasta li s-a acordat o atenţie specială.

Standardizarea din 1916 a fost făcută pe 1.700 de copii normali, 200 de copii deficienţi şi superior dotaţi şi 400 de adulţi, importanţă deosebită având selecţionarea cazurilor pentru constituirea de grupe semnificative din populaţia americană, acţiune de pionierat în dezvoltarea generală a eşantionării în construcţia de teste. Pentru fiecare test component erau oferite instrucţiuni amănunţite de administrare şi stabilire a scorurilor. Fiecare test component era un experiment prescurtat, elementele constitutive fiind observate, controlate în privinţa condiţiilor în care se produceau şi interpretate pe cât posibil obiectiv. Douăzeci de ani de folosire a scalei pentru cercetarea şi diagnosticul clinic au constituit o bază experimentală complexă pentru următoarea revizie. Diferitelor părţi ale scalei le-a fost stabilită validitatea prin împărţirea copiilor din fiecare nivel de vârstă în trei categorii: valori IQ peste 110, valori IQ între 90 şi 109 şi valori IQ sub 90. Validitatea şi fidelitatea erau satisfăcătoare pentru o gamă largă de domenii, dar, pentru partea inferioară a abilităţilor adulţilor, erau considerate inadecvate. Anumite teste componente aveau o slabă validitate, iar instrucţiunile în privinţa stabilirii scorurilor erau deficitare ca precizie, fapt ce afecta asigurarea obiectivităţii şi comparabilităţii rezultatelor. Revizia din 1937 cuprinde un scurt rezumat al procedurilor de standardizare şi reexaminare, din cauza unor curente de opinii critice care dovedeau neînţelegerea procedurilor ce trebuiau folosite în dezvoltarea scalei şi din cauza curentului popular care folosea IQ-ul ca sinonim pentru inteligenţă, expectanţă nerealistă asociată testelor de inteligenţă.

Fidelitatea şi validitatea. La forma din 1937, fidelitatea era calculată prin corelaţia formelor „L" cu formele „M", găsindu-se un înalt grad de relaţionare. Au existat anumite implicaţii clinice la relaţionarea între fidelitate şi vârstă, fidelitate şi nivelul IQ al subiecţilor. Analiza lui McNemar asupra standardizării datelor arată că scala era mai fidelă pentru copiii cu vârstă crescută decât pentru cei de vârstă mai mică.

De asemenea, s-a realizat calcularea fidelităţii pe vârste diferite şi în funcţie de magnitudinea IQ-ului. Astfel, de la 2,5 la 5,5 ani, coeficientul de fidelitate trecea de la 0,83 (pentru IQ 140-149) la 0,91 (pentru IQ 60-69). La vârsta de la 6 la 13 ani, se trecea de la 0,91 (pentru IQ 140-149) la 0,97 (pentru IQ 60-69). La vârsta de la 14 la 18 ani, trecerea era de la 0,95 (pentru IQ 140-149) la 0,98 (pentru IQ 60-69).

Variabilitatea scorurilor vârstei mintale creşte cu avansarea în vârstă. Se aştepta ca fluctuaţia subiecţilor în scorul vârstei mintale să nu fie proporţională cu variabilitatea vârstei cronologice a grupului, ci cu variabilitatea vârstei mintale.

Iată o scală în care ordonarea cantitativă se face în funcţie de valorile IQ-ului; la fiecare gradient este anexată o apreciere calitativă:

- peste 150 – aproape geniu;
- 130-149 – înalt superior;
- 115-129 – superior;
- 85-114 – normal;

- 70-84 - uşor deficient;
- 60-69 - intelect de limită;
- 40-59 - *moron*;
- 20-39 - imbecil;
- 0-19 - idiot (*apud* Roşca, 1972, p. 181, cu o uşoară „retuşare" a termenilor).

Redăm un tabel normativ în două viziuni: Stanford-Binet şi D. Wechsler[1] (exprimarea se face în procente):

Tabelul 1

Procentajul din populaţie al diferitelor niveluri de inteligenţă, exprimate în IQ (în viziunea lui Terman-Merrill şi D. Wechsler)

IQ	Denumire	Stanford-Binet	D. Wechsler
peste 129	Inteligenţă excepţională	4,4	2,2
120-129	Inteligenţă înaltă	8,2	6,7
110-119	Inteligenţă bună	18,1	16,1
90-109	Inteligenţă medie	46,5	50
80-89	Inteligenţă normală slabă fără anomalie propriu-zisă	15,5	16,1
mai puţin de 70	Debili	2,7	2

Pentru mai multă operativitate, autorii dau tabele cu coeficienţi corespunzători diferitelor vârste mintale şi cronologice. Utilitatea acestor tabele este resimţită mai ales pentru vârste de peste 13 ani, domeniu în care raportarea se face la o vârstă cronologică ajustată, şi nu la una reală. Revizuirea din 1960 introduce *coeficientul de inteligenţă de deviaţie*, IQD, care are media 100 şi abaterea standard 16 pentru toate grupele de vârstă. În tabelele de calculare a IQ-ului se trece de limita de 16 ani, ajungându-se la vârstele de 17 şi 18 ani.

Perfecţionarea atentă face ca scala creată de Terman şi Merrill să fie un reper principal pentru a judeca multitudinea de teste care au urmat construcţiei sale. Logica acestor noi teste era următoarea: „Dacă testul Stanford-Binet măsoară inteligenţa şi dacă testul nostru corelează puternic cu acesta, înseamnă că şi testul nou creat măsoară inteligenţa".

La puţin timp după lansarea versiunii Terman-Merrill, testele au făcut obiectul unei industrii în care erau mulţi dolari în joc. Societăţile care comercializau teste nu îndrăzneau să-şi asume riscul de a propune teste a căror validitate nu a fost verificată prin adecvarea lor la normele fixate de Terman. În faţa presiunii banilor şi a timpului, s-a renunţat la multe dintre exigenţele de minuţiozitate şi calm în aplicare. Chiar Terman face concesii, în momentul când îşi propune să examineze un număr cât mai mare de copii.

6) Planul eşantionării şi administrării. Revizia din anul 1916 se caracterizează prin aceea că „întinde" vârsta de aplicare la „adulţii superiori" şi face ca numărul probelor să crească la 90, de la 54, cât cuprindea ultima versiune a lui Binet, din anul 1911. Pentru

1. Vezi Peter R. Hofstätter (coord.), *Psychologie*, Fischer Taschenbuch Verlag, Frankfurt pe Main, 1971, p. 289.

a stabili clar eşantioane de vârstă, examinarea s-a făcut cu o toleranţă de două luni înainte şi după data naşterii. Această primă revizie a fost precedată de o masivă acumulare de date cu referire la rezultatele aplicării, în diferite ţări, a scalei Binet-Simon şi mai ales la distribuţia pe vârste a performanţelor. Standardele şi performanţele intelectuale sunt furnizate pentru copii născuţi în Statele Unite, având între 3 şi 16 ani. Abilitatea intelectuală individuală reflectată în performanţă individuală era apreciată prin comparaţie cu standardele de performanţă ale copiilor de aceeaşi vârstă.

Evaluarea se făcea în termeni de vârstă mintală. În urma examinării, o persoană putea fi considerată normală (când corespundea standardelor stabilite pentru aceeaşi vârstă), retardată (dacă rezultatele corespundeau standardelor pentru persoane mai tinere) sau avansată (dacă rezultatele corespundeau standardelor persoanelor mai în vârstă). Noţiunea de *vârstă mintală* se folosea din raţiuni de continuitate a terminologiei originale şi din cauza uzitării frecvente a acestui termen în domeniul clinic. Termenul de IQ este introdus, în varianta din 1916, ca fiind raportul între vârsta mintală şi vârsta cronologică.

Testele au fost grupate pe jumătăţi de an. Până la nivelul de 5 ani au fost păstrate testele alternative. La fiecare grup de vârstă era testat un grup reprezentativ. O mare atenţie a fost acordată volumului eşantionului de etalonare şi exigenţei ca el să fie reprezentativ. Totuşi, nivelul economic al celor testaţi a fost superior populaţiei generale, iar elementul urban a fost predominant. În 1937, grupul standardizat cuprindea 3.184 de copii albi, născuţi în Statele Unite, câte 100 de subiecţi la fiecare jumătate de an, de la 1,5 ani la 5,5 ani, câte 200 de subiecţi de la 6 la 14 ani, câte 100 de la 15 la 18 ani. Fiecare grup era egal divizat între sexe. Când diferenţele erau de jumătate de an, grupurile erau omogene, iar grupurile de vârstă sub 7 şi peste 14 ani creau probleme.

În privinţa distribuţiei geografice, se asigura un echilibru, la examinare participând 17 comunităţi din 11 state separate, din mediile urban, suburban şi rural. Se căuta ca grupele de vârstă să se obţină la întâmplare. Exista un număr disproporţionat de subiecţi din mediul urban, din cauza indicaţiei privind nivelul economic.

Ordinea administrării probelor este prevăzută în manualul scalei, dar ulterior examinatorului i se permite să aleagă ordinea probelor. Autorii pun accentul pe conformitatea răspunsurilor la previziuni şi penalizează orice dovadă de creativitate. Iată un caz relatat de Gould, referitor la adăugirile lui Terman faţă de versiunea originală a lui Binet: „Un indian, venit într-un oraş pentru prima dată în viaţa sa, vede un alb pe stradă şi exclamă: «El pare a fi un leneş de merge aşezat». Pe ce stătea albul pentru ca indianul să poată spune că el merge aşezat?" (vezi Gould, 1984, p. 217). Singurul răspuns corect pe care îl acceptă Terman este bicicleta, refuzând diferitele feluri de vehicule sau alte răspunsuri care ar fi arătat originalitate. Examinarea începe de la acel nivel care se presupune că va fi trecut în întregime de către subiect. Se continuă cu testele de la nivelurile următoare, până când nici un test nu mai este reuşit. „Scatter"-ul sau dispersia nereuşitelor are o semnificaţie diagnostică.

În general, pentru aplicare sunt necesare între 30 şi 40 de minute pentru vârstele mici şi 60-80 minute pentru vârstele mari. Observarea intervenţiei stării de oboseală poate determina aplicarea scalei în două etape. În situaţia de presiune temporală se poate aplica forma abreviată formată din 3-5 teste pentru fiecare nivel. Era indicat ca la această formă să nu se apeleze decât dacă erau disponibile o serie de date cu valoare psihodiagnostică compensatorie. Indicaţiile din manual privind cotarea sunt foarte amănunţite. De altfel, întregul manual este foarte minuţios şi nimic nu este lăsat la voia întâmplării.

Totuşi, standardizarea cotării nu este realizată deplin. Date complementare rezultate din observarea atentă a subiectului compensează această deficienţă, cu toate că noi considerăm că o anumită libertate de cotare a răspunsurilor este benefică şi chiar indicată. Calculul vârstei mintale se face prin adăugarea la vârsta mintală de bază (la care s-au efectuat toate testele) a valorilor parţiale obţinute. Terman şi Merrill avertizează că aceste vârste mintale au cota lor de arbitrar, mai ales când studiem o realitate foarte complexă aflată în evoluţie.

Studiul lui Terman asupra geniului. În rândul creatorilor de teste psihologice, s-a pus permanent problema aprecierii supra- şi subdotaţilor şi a necesităţii de a realiza studii longitudinale pentru a crea un fundament pentru teoriile naturii inteligenţei. Lewis Terman realizează în această privinţă studii excepţionale, care vor fi considerate multă vreme etalon în domeniu. În vremea lui Terman, era larg răspândită credinţa că elevii străluciţi erau anormali, mai mult decât în sens statistic. Termenul de *anormal* era lipsit, în contextul dat, de semnificaţia peiorativă.

Într-o expresie comună, se spunea că aceşti copii erau „mai devreme maturizaţi, mai devreme degradaţi", sugerând că, dacă abilităţile mintale se dezvoltau prea puternic la o vârstă fragedă, nu mai rămâne îndeajuns pentru anii de mai târziu. Prin studii obiective de-a lungul anilor cu copii supradotaţi, Terman a realizat unele dintre cele mai faimoase studii din istoria psihologiei.

Prin identificarea copiilor foarte inteligenţi şi observarea acestora de-a lungul unei perioade de timp, Terman poate să ilustreze convingerea proprie conform căreia copiii cu un IQ înalt aveau mai mult succes în viaţă decât cei cu un IQ slab. La acest prim pas, Terman îl considera geniu pe cel ce obţinea la testul său un scor de 135 sau mai mult. El şi colegii săi au administrat testul la mii de şcolari din California şi au identificat 1.528 de copii de vârf (dintre care 856 băieţi şi 672 fete). Media vârstei cronologice a grupului era de 11 ani şi media IQ-ului grupului era de 151. Inventariind tot ceea ce se putea de la aceşti subiecţi, incluzând interesele, istoria familiei, sănătatea, caracteristicile fizice şi personalitatea, Terman a încercat să studieze experienţa membrilor grupului, o dată cu maturizarea lor pe parcursul anilor. Îşi începe studiul în 1921 şi o primă raportare a rezultatelor este realizată în *Genetic Studies of Genius*, lucrare apărută în 1926 (detectarea a fost făcută în 1914). Cititorul este avertizat în privinţa dublului sens al termenul *genetic*. Un prim sens este acela de „dezvoltare în timp". Când termenul a fost utilizat cu această semnificaţie, un studiu genetic desemna unul care stabilea cât de mult variază o caracteristică psihologică în procesul de maturizare. În al doilea sens, termenul *genetic* poate să se refere la gena sau cromozomii responsabili pentru variate trăsături. Terman foloseşte termenul în sens de „dezvoltare în timp". El găseşte că şcolarii luaţi în studiu (denumirea acestora este de *Termite*, în sensul de „supradotaţi descoperiţi de Terman") au părinţi cu zestre intelectuală peste medie, au învăţat să citească de la vârste mici, au participat la o diversitate de activităţi şi au avut un parcurs şcolar de regulă excelent. Chestiunea majoră era câţi dintre aceşti copii vor prospera înaintând în vârstă.

Terman realizează un alt studiu longitudinal, în 1927-1928, când vârsta medie a grupului ajunge la 16 ani şi, ulterior, în perioada 1939-1940, când vârsta medie era de 29 de ani. Aceste studii indicau faptul că scorurile la test aparţineau părţii superioare de 1% din populaţia generală şi că membrii grupului participau la o mare varietate de activităţi, excelând la cele mai multe dintre acestea, deoarece erau realizaţi în plan

academic: 70% dintre bărbaţi şi 67% dintre femei au terminat colegiul, 56% dintre bărbaţi şi 33% dintre femei au atins un grad de realizare avansat. Toate aceste procente erau mult superioare celor din populaţia generală.

În 1947, Terman apare la un show radiofonic, *Quiz Kids*. Cu această ocazie, copiilor performeri, sănătoşi li se puneau întrebări foarte dificile, la care era foarte greu să se răspundă în mod obişnuit. Terman apărea în program, prezentându-se ca responsabil pentru corectarea câtorva dintre prejudecăţile despre copiii dotaţi. După moartea lui Terman, s-a aflat că răspunsurile erau cunoscute dinainte.

Finalul studiului longitudinal, cu participarea lui Terman, a fost realizat în 1950-1952 şi arată că membrii grupului continuau să exceleze în cele mai multe dintre categoriile studiate. S-a observat că unii membri ai grupului s-au afirmat ca doctori, avocaţi, profesori, judecători, ingineri, scriitori, actori, oameni de ştiinţă şi conducători politici. După moartea lui Terman, în 1956, coordonarea investigaţiei a fost preluată de Robert R. Sears, profesor la Stanford, care a făcut parte dintre *Termite*. În 1970, alţi doi profesori de la Stanford au fost cooptaţi în echipa de investigaţie: Lee J. Cronbach, şi el dintre *Termite*, şi Pauline S. Sears, soţia lui Robert.

Cea mai recentă fază de colectare a datelor studiului a constat în completarea din 1956, sub îndrumarea lui Robert Sears şi Albert Hastorf.

Grupul vârfurilor individuale, identificat de Terman în 1914, a fost studiat intens mai mult de 60 de ani. Studiile continuă chiar după ce o serie de *Termite* s-au stins din viaţă. De exemplu, Tomlinson-Keasey şi Little (1990) au examinat realizările a 1.069 de *Termite* din lotul original de 1.528 şi au conchis că persoanele studiate au avut mare succes şi s-au adaptat mai bine. Tomlinson-Keasey şi Little au izolat variabilele referitoare la achiziţiile diferenţiale şi la nivelul de adaptare, aşa încât acestea puteau să fie folosite pentru predicţie şi pentru a mări achiziţiile şi adaptările altor supradotaţi individuali. Friedman, Tucker, Schwartz, Tomlinson-Keasey, Wingard şi Criqui (1995) au examinat fundamentele grupului *Termitelor* lui Terman, care nu mai erau în viaţă din 1991. Ei au descoperit că anumite variabile psihosociale comportamentale erau predictori semnificativi ai mortalităţii premature, ca, de exemplu: divorţul parental în perioada copilăriei, *pattern*-ul de mariaj instabil care se derula în perioada adultă, anumite caracteristici ale personalităţii în copilărie, instabilitatea psihică în perioada adultă şi obiceiurile nesănătoase (fumatul şi consumul excesiv de băuturi alcoolice)[1].

Pentru cercetătorii implicaţi în studiul longitudinal al lui Terman, rezultatele primare erau clare: copiii supradotaţi se adaptează la adulţii supradotaţi. Studiile lui Terman înlătură câteva prejudecăţi despre copiii supradotaţi, dar rămâne fără răspuns întrebarea: câtă supradotare este moştenită şi câtă supradotare se datorează experienţei? Terman avea credinţa că supradotarea era moştenită, dar, ulterior, cercetările au arătat că unele dintre rezultatele lui Terman pot fi considerate ca fiind influenţate de experienţa membrilor grupului. Cât de mult din inteligenţă este genetic determinată şi cât de mult este influenţată ambiental – aceasta va constitui o controversă fierbinte în psihologie. Mai multe cercetări moderne conchid că ambii factori prezintă importanţă.

1. Referinţele statistice sunt preluate din R. Ernest Hilgard, *Psychology in America. A Historical Survey*, Harcourt Brace Jovanovich, Publishers, San Diego, New York, Chicago, Austin, 1987, pp. 463-466, şi B.R. Hergenhahn (1997), *An Introduction to the History of Psychology* (ediţia a III-a), Brooks/Cole Publishing Company, Pacific Grove, Albany, pp. 287-288.

În câteva cazuri, studiile longitudinale ale individualităţilor supradotate arată clar că acestea au obţinut scoruri înalte la probele psihologice, care îşi propun să măsoare inteligenţa precoce, scoruri ce rămân constante mai târziu în viaţă. În fapt, rezultatele lor arată că cei care fac bine şi corect un lucru în tinereţe tind să facă la fel şi la vârsta adultă.

4.1.3. *Reacţia faţă de ereditarism*

Pentru Terman, primul beneficiu al testării mintale era identificarea indivizilor supradotaţi, aşa încât aceştia să poată fi încurajaţi să înveţe la înaltul lor potenţial şi să ajungă lideri sociali. El credea că un sistem de urmărire, ca acela al studenţilor supradotaţi, care a furnizat experienţe educaţionale diferite de ale studenţilor obişnuiţi, era esenţial în supravieţuirea democraţiei. În special, dintre eforturile lui Terman şi ale colegilor săi, testarea psihologică şi gruparea abilităţilor erau practici comune în şcolile elementare din SUA, în 1930. Deşi recomanda intens un curriculum şcolar diferenţiat, Terman nu a făcut recomandări speciale cuprinzând metode ale educaţiei adaptate nevoilor copiilor cu intelect superior.

Leta Stetter Hollingsworth (1886-1939) a fost primul cercetător care s-a ocupat de strategii de dezvoltare educaţională, menite să asigure buna dezvoltare a studenţilor supradotaţi.
Născută Stetter, obţine *bachelor's degree* la University of Nebraska. În 1908, pleacă la New York, unde este angajată ca psiholog instructor la Barnard College. Leta Hollingsworth intenţiona să rămână în învăţământ la New York, însă descoperă că oraşul urma politica de a nu angaja femei ca profesori. Se înscrie ca *graduate student* la Columbia University, unde urmează cursurile lui Edward L. Thorndike, căruia îi devine consultant. Luând contact cu ambianţa cercetărilor lui Thorndike, a devenit foarte preocupată de testarea psihologică. O interesau în mod special prejudecăţile faţă de femei, care erau predominante în acel timp. Deşi Leta Holingsworth nu se aştepta, Thorndike agreează să-i îndrume disertaţia *Functional Periodicity*, pe marginea ipotezei potrivit căreia femeia era perturbată în perioada de menstruaţie. Nu găseşte dovezi pertinente pentru o astfel de perturbare (1914).
După obţinerea masteratului, în 1917, Leta Stetter Hollingsworth lucrează pentru o perioadă de timp ca psiholog clinician la New York City Clearing House for Mintal Defectives, unde administrează scala Binet-Simon. Lucrează şi la Bellevue Hospital ca psiholog clinician, perioadă în care reuşeşte să obţină doctoratul la Columbia University, în 1916. Primeşte apoi postul de profesor de educaţie la Teachers College, Columbia University. Munca la Clearing House o face să scoată în evidenţă câteva mituri despre aşa-zisa mentalitate defectivă individuală cu privire la femei. Descoperă astfel că mulţi indivizi erau clasificaţi drept „defectivi", în realitate fiind vorba despre o manifestare locală şi o problemă de adaptare personală. Scrie o serie de cărţi, în care îşi propune să corecteze această percepţie : *The Psychology of Subnormal Children* (1920), *Special Talents and Defects : Their Significance for Education* (1923), *The Psychology of the Adolescent* (1928), recomandate de G. Stanley Hall ca standarde în domeniu.

Leta Stetter Hollingsworth contestă credinţa larg acceptată că inteligenţa era, în mare măsură, moştenită şi că femeile sunt inferioare bărbaţilor. La acea vreme, Thorndike se numără printre aceia care nu împărtăşeau astfel de credinţe. Leta consideră că femeia deţine mai puţine poziţii proeminente decât bărbatul, nu din cauza inferiorităţii intelectuale, ci din cauza rolurilor sociale care îi revin cu precădere.

După colaborarea cu Leta Stetter Hollingsworth, Thorndike îşi modifică viziunea asupra inteligenţei, referitoare la aportul superior al educaţiei faţă de cel al naturii. Leta Stetter Hollingsworth considera că este parţial responsabilă de schimbarea acestei concepţii. Ea discută şi cu Terman despre convingerile sale potrivit cărora mai mulţi bărbaţi decât femei sunt clasificaţi ca supradotaţi nu din cauza diferenţierii abilităţilor intelectuale, ci din cauza factorilor sociali. Terman face şi el o serie de modificări în concepţiile lui nativiste, care se refereau la generarea diferenţelor în inteligenţă de către influenţele sociale, însă menţine fondul convingerilor sale după care inteligenţa era, în primul rând, determinată genetic.

Leta Stetter Hollingsworth îşi concentrează apoi atenţia asupra copiilor supradotaţi. Ea susţine că simpla clasificare a copiilor ca supradotaţi nu era suficientă. Prin accentuarea scorului abstract la test sau a caracteristicilor de grup, nevoile individuale ale celor ce studiază erau scăpate din vedere. Se descrie experienţa unei fete supradotate de 8 ani, numită Jean, care termina mult mai repede sarcinile decât ceilalţi elevi din clasa sa. Reacţia profesorului la această problemă era să o pună pe Jean să scrie numere într-un caiet până ce restul clasei termina tema încredinţată.

Corectarea tratamentului inadecvat acordat supradotaţilor devine preocuparea principală a acestei cercetătoare pasionate, pentru tot restul carierei sale. În 1926, publică *Gifted Children*, care conţine teste standard în scalele de educaţie pentru mulţi ani, şi *Children Above 180 IQ*, publicată postum, în 1942.

4.1.4. *Gigantismul în testarea psihologică*

Robert M. Yerkes este una dintre cele mai proeminente figuri din întreaga istorie a psihologiei americane. Ca particularităţi ale personalităţii sale, amintim faptul că avea o fizionomie de frustrat, avea aptitudini de conducere, era un înfocat şi elocvent promotor al profesiei sale. A trăit în perioada în care psihologia nu era o ştiinţă apreciată în mediile academice, în unele cazuri chiar contestându-i-se statutul de ştiinţă. Yerkes dorea să contribuie la recunoaşterea unui statut propriu pentru disciplina pe care o profesa cu multă rigurozitate. Ca majoritatea contemporanilor săi, el asimila rigoarea şi ştiinţa cifrelor şi cuantificării. Exprimată în mod plastic, marea sa ambiţie a fost de a transforma psihologia dintr-o artă contestată într-o ştiinţă respectată. Considera că testele mintale constituiau o sursă promiţătoare de cifre abundente şi obiective. În concepţia sa, psihologia putea cuceri poziţia de ştiinţă veritabilă, susceptibilă de sprijin financiar şi instituţional, doar dacă ea reuşea să plaseze chestiunea potenţialităţilor umane sub auspiciile unei metodologii foarte riguroase. În vremea în care Yerkes se afirmă în mişcarea testelor psihologice, acestea sufereau de o lipsă de mijloace, dar şi de propriile lor contradicţii. Ele erau practicate de amatori rău formaţi, ale căror rezultate evident absurde dădeau un fals renume acestei tehnici. În 1915, la reuniunea American Psychological Association, se raporta că primarul din Chicago, testat cu o versiune a scalei Binet-Simon, a obţinut notele unui debil. De asemenea, alte scale disponibile în acea perioadă

ofereau note ce duceau la concluzii divergente, chiar în situaţia în care ele erau folosite corect. De exemplu, notele slabe la scala Stanford-Binet, care indicau totuşi zona normalului, desemnau un *moron* când erau rezultate din aplicarea unei versiuni Binet concepute de Goddard. Indicatorii statistici lipseau, iar cooperarea între specialişti a fost prea sporadică pentru ca o persoană să poată constitui un rezervor de date suficient de bogate şi omogene pentru a aduce o anumită clarificare.

Robert Mearns Yerkes (1876-1956) a fost primul născut al unei familii de fermieri din Pennsylvania. Ca şi Terman, face parte dintre cei care au pornit „de jos". Este dezamăgit de viaţa de la fermă şi aspiră să devină medic. Se înscrie la Ursinus College pe care îl termină după ce cunoaşte experienţa efortului de a se ajuta singur. După Ursinus, absolvit în 1897 cu un *BA* (*bachelor of Arts*), Yerkes se înscrie la Harvard College, unde reuşeşte de asemenea un BA, în 1898. Rămâne un an tot la Harvard, ca *graduate student* în zoologie, interesându-se de comportamentul animalelor, fapt ce are drept rezultat publicarea unei prime lucrări ştiinţifice pe această temă. Ca urmare a influenţelor venite din partea lui Hugo Münsterberg şi al lui Josiah Royce, Yerkes îşi îndreaptă preocupările spre domeniul psihologiei şi obţine doctoratul în 1902 cu o teză despre reacţiile senzoriale la meduză. Rămâne la Harvard timp de 15 ani: între 1902 şi 1908 este instructor, iar între 1908 şi 1917 funcţionează ca *assistant professor* la psihologie comparată. Împreună cu prietenul său, John B. Watson, care activa la Johns Hopkins University, Yerkes fondează psihologia comparată în SUA. Împreună cu E.B. Holt, proiectează şi echipează în şapte camere Laboratorul de psihologie comparată, inaugurat la Harvard în 1905. Deoarece, ca student, Yerkes a împrumutat sume considerabile de bani şi pentru că postul de la facultatea din Harvard nu-i aducea suficienţi bani, este nevoit să lucreze în mai multe locuri pentru a face faţă acestei situaţii. În 1912, este director de cercetări psihologice la Boston State Psychopathic Hospital. Aici face prima experienţă cu testele psihologice şi utilizează scala Binet-Simon ca instrument de ajutor diagnostic. Unul dintre profesorii lui Yerkes la Harvard, ulterior prietenul şi colegul său, era biologul Charles Davenport, care coresponda cu Galton şi era lider al mişcării eugeniste în SUA. Yerkes devine un puternic avocat al eugenismului. Se adaptează evoluţiilor testării la Boston Psychopathic Hospital şi realizează cercetări în domeniul psihologiei comparate. Descurajat de ritmul foarte lent al avansării profesionale la Harvard, Rober Yerkes acceptă, în anul 1917, invitaţia de a conduce Departamentul de Psihologie al Universităţii Minessota. La cele două universităţi americane cunoaşte tineri psihologi, care devin colaboratori la construcţia scalelor de inteligenţă folosite în primul război mondial (J.W. Bridge, R.S. Harwick, R.M. Elliot şi Karl S. Lashley). Ocupă în continuare funcţii în organisme ştiinţifice americane de nivel naţional şi întreprinde cercetări de anvergură în domeniul psihologiei comparate. Interesul nostru, în această lucrare, este legat de contribuţia sa la metodologia creării şi aplicării testelor psihologice.

Aportul lui Yerkes la metodologia de testare a inteligenţei a constat în primul rând în sugestia potrivit căreia examinaţii să efectueze toţi itemii la testul Binet-Simon şi să se acorde puncte pentru fiecare item rezolvat corect. Aduce argumente în favoarea opiniei că scorul obţinut de o persoană trebuie să fie exprimat în termeni de *puncte total câştigate*, în loc de IQ. Acest mod de a proceda înlătură vârsta ca factor de scorare. Procedura tradiţională, urmată în administrarea scalei Binet-Simon, viza adaptarea testului pentru o caracteristică individuală. De exemplu, dacă un copil în vârstă de 7 ani era testat, urma să fie administrat testul apropiat de această vârstă. Dacă el greşea ceva din test sau testul în întregime, urma să fie administrat testul adaptat unei vârste mai mici. Dacă însă copilul rezolva iniţial toate probele adaptate pentru vârsta de 7 ani, urma să fie administrat nivelul de testare pentru nivelul de 8 ani ş.a.m.d., până când copilul începea să greşească. Altfel spus, prin folosirea vârstei ca un cadru de referinţă, procedura de testare era adaptată pentru fiecare copil. „Scala de puncte" a lui Yerkes făcea toate aceste proceduri inutile.

Yerkes considera că, dacă punctele normei puteau fi stabilite pentru vârste variate şi pentru fiecare grup pe care urma să fie comparat, testul era mai uşor de administrat şi punctajele testelor erau mai adaptate la analiza statistică, decât scorurile IQ. Din cauza scorului realizat la o serie de teste privite prin prisma vârstei sau a nivelului, metoda lui Yerkes conduce la testarea de grup.

Pe când se afla la Boston Psychopathic Hospital, Robert Yerkes, împreună cu J.W. Bridge, încearcă să reformeze probele lui Binet, clasându-le după procesele mintale la care ele fac apel: receptivitate, raţionament, perceptivitate, afectivitate, imaginaţie, memorie, gândire, emoţie, judecată etc. Ei constituie o serie de măsurători pentru vârstele plecând de la 2 ani şi încearcă să obţină o gradare în dificultate pentru fiecare test sau fiecare parte a testului în aşa fel încât măsurarea să se facă la fel de uşor şi precis pentru copil şi adult, utilizând pentru această operaţiune o serie de cotaţii particulare. El grupează 50 de puncte pentru fiecare dintre marile sale categorii şi obţine rezultate de forma următoare:

$$\mathrm{IQ} = \frac{r}{R} + \frac{i}{I} + \frac{a}{A} + \frac{p}{P},$$

în care R, I, A, P sunt rezultatele obţinute de un subiect mediu de vârstă mintală considerată şi r, i, a, p rezultatele subiectului. La numărător se află deci nota obţinută de subiect, iar la numitor nota medie a vârstei subiectului. Coeficientul obţinut variază între 0,91 şi 1,10.

Meritele sale în cercetarea şi metodologia psihologică determină alegerea lui Yerkes în funcţia de preşedinte al American Psychological Association (APA), în 1917.

Programul de testare a armatei. La apropierea mobilizării pentru primul război mondial, Yerkes a avut o „mare idee", dintre acelea care fac ca o persoană să rămână în istoria ştiinţei, şi anume de a examina psihologic majoritatea militarilor care participau la efortul de război al Statelor Unite ale Americii. De altfel, în majoritatea statelor europene angajate în conflict, psihologii contribuie, cel mai adesea prin organizarea acţiunilor de selecţie, la efortul de război: Myers o face pentru armata britanică, Moede – pentru armata germană (selecţia şoferilor), Lahy – pentru armata franceză şi Gemelli – pentru armata italiană (selecţia aviatorilor). Dar, în SUA, mobilizarea psihologilor va lua o cu totul altă amploare. Pentru Robert Yerkes, această masivă examinare era o excelentă ocazie de a acumula o mare cantitate de cifre, omogene şi practice, procedură care, pentru psihologie, reprezenta singurul drum ce îi putea asigura statutul de ştiinţă adevărată. El reuşeşte să convingă comandanţii militari, cu excepţia celor de la marină, de utilitatea unui examen psihologic al masei militarilor recrutaţi. Yerkes era convins că psihologii

pot să contribuie la selectarea şi evaluarea recruţilor din forţele armate. Ca urmare a activităţii sale în domeniul militar, primeşte gradul de maior în armata americană, motiv pentru care în unele cazuri este desemnat ca *maiorul Yerkes* (vezi Hilgard, 1978, p. 190). Acţiunea era de mare amploare : practic, se urmărea integrarea a 3,5 milioane de oameni într-o structură militară care cuprindea 500.000 de oameni din armata regulată. Pentru a putea duce la bun sfârşit o astfel de sarcină, Yerkes a convocat o adunare a psihologilor, pentru a stabili câţi dintre aceştia doreau să participe la efortul de război. Convocarea nu a trezit mare entuziasm la început şi, la invitaţia lui Goddard, un mic grup de psihologi, incluzându-i pe Terman şi Yerkes, a mers la Vineland Training School pentru a dezvolta testele psihologice, care au fost aplicate ulterior la numeroase trupe şi baze navale. Cum rezultatele erau încurajatoare, lui Yerkes i s-a dat sarcina de a organiza programul de testare psihologică pentru întreaga armată (marina a respins cererea). Pentru a elabora testele care urmau să fie aplicate, Yerkes a făcut apel la principalii ereditarişti din psihologia americană. Din mai până în iulie 1917, a lucrat alături de Lewis Terman, Henry Goddard şi alţi colegi la Vineland Training School din New Jersey, instituţie pe care o conducea părintele moronilor. După conturarea iniţiativei, psihologii au răspuns în număr foarte mare. În aprilie 1917, sunt create mai multe comitete de psihologi pentru examenul recruţilor. Un comitet a fost însărcinat cu recrutarea piloţilor, utilizând un test pus la punct de E.L. Thorndike. Un alt comitet (comitetul de clasificare a personalului), condus de Dill Scott, evalua, cu ajutorul unei scale de evaluare, şi nu al testelor, competenţele profesionale ale recruţilor. Apoi, comitetului pentru examinarea psihologică a recruţilor, coordonat de R. Yerkes, i se încredinţează o sarcină imensă : testarea ansamblului militarilor.

Scopul acestui program era de a-i identifica pe acei recruţi care prezentau deficienţe mintale, de a clasifica militarii în funcţie de nivelul de inteligenţă şi de a selecta persoane pentru un antrenament special, de exemplu pe cei care vor deveni ofiţeri. Yerkes credea că, pentru a fi funcţional, testul folosit pentru recruţi trebuia să fie mai degrabă un test de grup decât unul individual, care ar fi trebuit să măsoare inteligenţa nativă şi care ar fi fost uşor de administrat şi scorat. Folosind metoda lui Yerkes de scorare, grupul de psihologi lucrează la un test ce satisfăcea aceste criterii, dar a descoperit că aproximativ 60% dintre recruţi nu ştiu să citească bine. Grupul a rezolvat problema prin crearea a două forme de teste concepute pentru a fi aplicate colectiv : *Army Alpha*, pentru cei alfabetizaţi, şi un test nonverbal, *Army Beta*, pentru indivizi analfabeţi, pentru cei care vorbeau curent altă limbă decât limba engleză şi pentru cei care ratau testul verbal. În caz de eşec la testul *Army Beta*, recruţii treceau un test de inteligenţă individual, cel mai adesea o versiune a scalei lui Binet.

În funcţie de rezultate, soldaţii erau clasaţi în cinci categorii, cărora le corespundeau diferite repartizări militare. Psihologii din armată clasau toţi recruţii pe o scală valorică de la A la E şi propuneau distribuirea militară adecvată. Cei de la nivelul C trebuiau să primească menţiunea „inteligenţă medie slabă" şi să devină simpli soldaţi. Cei clasaţi în D erau rar consideraţi apţi de a îndeplini sarcini necesitând o competenţă specială, simţul previziunii, al capacităţii de a se descurca sau o atenţie susţinută. Subiecţii clasaţi în treptele D şi E ale scalei nu puteau citi şi înţelege.

Descrierea testelor. Testele *Army Alpha* cuprindeau opt secţiuni, iar *Army Beta* – şapte. Efectuarea fiecărui test dura aproximativ o oră şi ambele puteau fi aplicate unor grupuri importante de subiecţi. Majoritatea secţiunilor prezentau tipuri de solicitări care au

devenit ulterior familiare generaţiilor de candidaţi la testele de tip analogie, suită de numere de completat, fraze în dezordine etc. *Army Alfa* este considerat modelul primar al tuturor tipurilor de solicitare exploatate în testele psihologice. Unele secţiuni sunt sensibile la influenţele culturale. Ele testează mai mult nivelul de alfabetizare şi de familiarizare cu mediul cultural american, decât inteligenţa înnăscută. În privinţa dificultăţii întrebărilor, ne putem face o impresie din componente de genul :

- *Crisco* este : o specialitate farmaceutică, un dezinfectant, o pastă de dinţi, un produs alimentar ?
- *Charles Mathewson* este celebru ca : scriitor, artist, jucător de baseball, comediant ?

Ne putem uşor imagina cam ce atitudine luau imigranţii recenţi faţă de asemenea întrebări. Timpul la teste era strict limitat, deoarece urma altă serie de subiecţi care aşteptau să fie examinaţi. Dificultăţile de explicare a probei, datorate presiunii timpului şi capacităţii examinatorilor, se adaugă la cele expuse până acum.

Condiţiile de realizare a examenului cereau ca acesta să se desfăşoare rapid, iar răspunsul să se coteze imediat în aşa fel încât cei care au eşuat să poată relua examenul. În multe cazuri, examinatorii erau obligaţi la nenumărate compromisuri, renunţări şi modificări. Procedurile, care variau de la o cazarmă la alta, abia puteau fi corelate şi comparate. Statisticile finale devin o armă în mâna eugeniştilor şi rasiştilor.

În privinţa condiţiilor de examinare, se poate preciza că armata a cerut clădiri şi instalaţii speciale pentru operaţiunea de testare, dar nimic nou nu s-a realizat, examinatorii confruntându-se cu lipsa unui minimum necesar.

Metodele de selecţie pentru *Army Beta* erau diferite de la o cazarmă la alta şi, uneori, chiar de la o săptămână la alta în aceeaşi cazarmă. Examenul *Beta* nu cuprindea decât cifre şi simboluri. Instrucţiunile cuprindeau amănunţit procedura de examinare, inclusiv coregrafia gestuală a examinatorilor şi asistenţilor.

Examinatorii s-au confruntat şi cu situaţii neplăcute de colaborare cu şefii militari. Cei din corpul de psihologi constituit de Yerkes au trebuit să înfrunte ostilitatea unei părţi din corpul profesionist al armatei. În unele cazuri, chiar înalţii comandanţi ai armatei au alcătuit comisii de anchetă asupra programului de testare.

Testele au exercitat o puternică influenţă în anumite sectoare, mai ales în cel care se ocupa cu selecţia militarilor ce doreau să intre în şcolile de ofiţeri. La începutul războiului, Garda Naţională număra 9.000 de ofiţeri, iar la sfârşitul acestuia număra 200.000, dintre care două treimi au ieşit din câmpul de instrucţie (unde au trebuit să treacă testele psihologice). Exigenţa de selecţie prevedea că nici o persoană inferioară nivelului C (adică D şi E) nu putea să intre în şcolile de ofiţeri. Probabil că ostilitatea ofiţerilor era influenţată şi de cei care erau până atunci selecţioneri şi care foloseau alte metode. Majoritatea *staff*-ului militar nu a fost convins de eficacitatea examinărilor, fiind semnalate numeroase conflicte între testatori şi acesta.

Războiul s-a încheiat în anul 1918 iar programul de testare s-a terminat în 1919, perioadă în care au fost testaţi peste 1,75 milioane de militari. Ne aflăm deja într-un domeniu al gigantismului psihometric. O serie de psihologi au emis ideea că programul de testare a armatei a demonstrat practicabilitatea psihologiei, dar evidenţele nu susţin o asemenea afirmaţie. Robert Yerkes va afirma mai târziu că testele psihometrice au contribuit la câştigarea războiului. Un studiu realizat în 1977 de Samelson pe această temă a stabilit că doar 0,005% dintre cei testaţi erau recomandaţi pentru concediere, ca nepotriviţi mintal. În câteva cazuri, cadrele armatei au ignorat recomandările lor.

Întrebarea care se pune este următoarea: dacă armata ar fi perceput programul de testare ca eficient, ea ar fi încheiat programul atât de repede după încetarea războiului? Testarea la scară mare a făcut ca testele să fie cunoscute de către un public larg şi a dat un impuls considerabil mişcării testelor. Acestea au devenit populare în Statele Unite şi numeroşi responsabili ai administraţiei şi ai întreprinderilor au dorit să le utilizeze pentru recrutarea şi pentru gestiunea personalului lor. După război s-a creat, în 1919, prima societate de consultanţă, Scott Company, de la numele fondatorului său, care utilizează serviciile unor psihologi de renume pentru a trata probleme de selecţie profesională. Tot imediat după război, se înfiinţează Psychological Corporation, primul organism eficient de comercializare a testelor (crearea unei astfel de corporaţii a fost încercată, fără succes, şi de James McKeen Cattell).

Army Alpha şi *Army Beta* au fost primele teste de inteligenţă aplicate pe scară largă. Într-o voluminoasă monografie, cuprinzând aproape 900 de pagini şi editată în 1921, cu titlul *Psychological Examining in the United States Army*, Robert Yerkes vorbeşte despre o mare cantitate de cereri venite din întreprinderi, instituţii şcolare şi din partea persoanelor particulare dornice de a utiliza metodele de examen folosite în armată şi de a le adapta propriilor nevoi. În Statele Unite ale Americii începea o perioadă a examinării psihometrice de masă.

Analiza rezultatelor. Influenţa primordială exercitată de teste rezultă din darea de seamă statistică urmând examinării propriu-zise, publicată de Robert Yerkes în 1921. Unul dintre cei mai cunoscuţi psihologi din perioada interbelică, Edwin G. Boring, care era locotenentul lui Yerkes în timpul examinării psihologice, a selecţionat 160.000 de dosare şi a extras din date un puternic halo ereditarist. Eşantionul era, într-adevăr, enorm şi presupunea stabilirea unor medii de ansamblu, medii naţionale şi rasiale, pentru subiecţi care erau în proporţii foarte diferite (negrii, de exemplu, erau într-o proporţie mai redusă decât albii). Din oceanul de cifre extrase de Boring, care au dăinuit şi au influenţat politica socială a SUA, rezultă următoarele concluzii:

- Vârsta mintală medie a americanilor albi se situa deasupra limitei debilităţii uşoare, la nivelul, înspăimântător de mediocru pentru ereditarişti, de 13 ani. În studii anterioare, Lewis Terman plasase linia de departajare la 16 ani. Această nouă cifră devine punctul de intervenţie pentru eugenişti, care prezic ruina ţării şi se lamentează cu privire la declinul inteligenţei americane, cauzat de reproducţia masivă a săracilor şi a debililor mintal, cu privire la diluarea inteligenţei filonului local prin propagarea sângelui negru prin metisaj şi cu privire la subminarea acestuia prin imigraţia avându-şi sursa în statele din sudul şi estul Europei. Robert Yerkes îşi dă seama de absurditatea afirmaţiilor şi susţine că cifrele se sprijină pe un eşantion de 62 de persoane (32 de elevi şi 30 de oameni de afaceri cu un nivel de educaţie limitat).
- Emigranţii europeni sunt clasaţi în funcţie de ţările de origine. Omul mediu din numeroase naţiuni este un debil. Popoarele înapoiate din Europa de Sud şi slavii din Europa de Est sunt mai puţin inteligenţi decât popoarele din Europa de Nord şi decât cele din Europa de Vest. Supremaţia inteligenţei nordice nu este o prejudecată şovină, susţineau autorii rapoartelor. Rusul mediu are o vârstă mintală de 11,34 ani, italianul de 11,01 ani şi polonezul de 10,74 ani.
- Negrii se situează la baza scalei, cu o vârstă mintală de 10,41 ani. În anumite cazărmi, s-a încercat a se diferenţia mai mult analiza pielii, în direcţii dictate de rasism. În cadrul analizei psihologice realizate la Camp Lee, negrii au fost împărţiţi în trei grupe, după intensitatea culorii închise a pielii lor. Grupul cel mai *alb* a obţinut rezultatul cel mai bun. Opiniile ofiţerilor se racordau la aceste cifre.

În cursul examinării, Yerkes a verificat prin teste mai multe prejudecăţi sociale, una dintre ele fiind aceea potrivit căreia majoritatea delincvenţilor provin din debili mintal. Apoi, psihologii lui Yerkes trec la examenul prostituatelor şi stabilesc că între 30% şi 60% dintre acestea sunt deficiente mintal.

Nici un rând dintre cele peste 800 de pagini ale raportului nu a fost scris pentru a pomeni de rolul pe care l-a jucat mediul în obţinerea unor astfel de rezultate. Ereditariştii, care au conceput sistemul de examinare, au rămas convinşi că ei au examinat inteligenţa înnăscută. Într-o circulară venită la şcoala de psihologie militară de la Camp Grenleaf se spune clar: „Aceste teste nu măsoară aptitudinile profesionale sau nivelul şcolar; ele măsoară capacitatea intelectuală. Aceasta se dovedeşte foarte importantă pentru evaluarea valorii militare" (Gould, 1984, p. 242). Trecerea anilor a produs modificări în concepţia lui Boring. Într-un studiu publicat în 1952, el vorbeşte de *deformări inconştiente ale rezultatelor* şi de *natura socială a adevărului*, considerând că „adevărul absolut nu este de dorit în ştiinţă" (Boring, 1963, cap. „The Validation of Scientific Relief", pp. 245-252).

Contradicţiile interne şi prejudecăţile sociale au invalidat total concluziile ereditariste ale lui Yerkes. Boring a calificat mai târziu aceste concluzii drept „groteşti". Condiţiile inadecvate de examinare făceau complet ridicole pretenţiile ca recruţii să fi putut fi într-o stare în care să-şi fi dezvăluit capacităţile înnăscute, cu toate că examinatorii erau instruiţi să inducă o atmosferă cordială. Pentru *Beta Army*, se impusese limitarea strictă a vocabularului. Amintim că, pe lângă examinator, mai era prezent la examinare şi un demonstrator, care urma instrucţiunile celui dintâi privind efectuarea testului. Condiţiile de examen şi chiar caracterul testului făceau ridicole pretenţiile de măsurare cu testul *Beta* a unei anumite stări care să se cheme inteligenţă. Aproape întotdeauna, examenul era condus în mod precipitat. Repetiţiile ulterioare pe studenţi au demonstrat că nici aceştia nu reuşeau să termine proba. Lipsa de relevanţă a examenelor se găseşte în statisticile finale, cărora Yerkes şi Boring le-au dat o interpretare diferită. S-au publicat, de asemenea, curbele de distribuţie a frecvenţelor pentru fiecare probă. Yerkes gândea că inteligenţa înnăscută era normal distribuită, după curba clasică a lui Gauss-Laplace, însă doar două probe, labirintul şi lacunele (1 şi 6), aveau o distribuţie apropiată de exigenţele teoretice ale celei *normale*. Toate celelalte distribuţii erau bimodale, cu evidenţieri ale unui punct de valoare medie şi ale unui punct de valoare minimală, adică 0. Bimodalitatea unei distribuţii empirice semnifică faptul că subiecţii aveau două feluri de răspunsuri faţă de test. Unii au înţeles instrucţiunile şi au răspuns în mod diversificat, alţii nu au înţeles instrucţiunile şi au obţinut 0. În interpretare statistică, multimodalitatea este indicele eterogenităţii în provenienţa valorilor unei variabile, situaţie care sugerează că fiecărei dominante îi corespunde o cauză distinctă. Aceste multimodalităţi ar fi trebuit să-i convingă pe Yerkes şi Boring că testele lor nu puteau să măsoare o entitate unică.

Problema valorii 0. Atât testul *Alpha*, cât şi testul *Beta* au avut probleme deosebite din cauza proporţiei mari a valorii 0. Dacă la testul *Beta* valoarea 0 nu reprezenta dominanta primară, la testul *Alfa* modulul avea această valoare. În unele teste, rezultatele nule atingeau 40% dintre valori. În mod normal, prezenţa masivă a valorii 0 demonstra clar că o serie de candidaţi nu au înţeles proba şi era necesară reluarea acesteia. Autorii au redus dificultatea primilor itemi de la mai multe probe, în scopul de a micşora numărul

de note nule. Boring a rezolvat problema fie excluzând valoarea 0 din statisticile finale, fie transformând, printr-un artificiu statistic, valoarea 0 în valori negative. Boring stabileşte o distincţie între „0 matematic", un minimum absolut sub care logic era imposibil să se poată coborî, şi un „zero psihologic", început arbitrar fixat la un test dat. Ca atare, el a găsit câteva posibilităţi de ajustare.

Monografia lui Yerkes furnizează o bogăţie de informaţii pentru acei care vor să cerceteze corelaţiile de ambianţă ale perfomanţelor în testele de inteligenţă. El a găsit o corelaţie puternică între rezultatele medii la teste şi infestarea cu ankilostom. În urma acestei constatări, a tras concluzia că starea de sănătate particular legată de sărăcie are efecte asupra notelor obţinute la teste. De asemenea, s-a stabilit o legătură între inteligenţă şi gradul de şcolarizare. El a găsit o corelaţie de 0,75 între rezultatele la teste şi numărul anilor de instrucţie. Un alt psiholog din echipa lui Yerkes, Montagu, a găsit puternice corelaţii între mărimea bugetelor de educaţie ale unui stat şi rezultatele medii ale recruţilor. Robert Yerkes identifică diferenţe considerabile între englezi, scandinavi şi germani, pe de o parte, şi latini şi slavi, pe de altă parte, dar recunoaşte că majoritatea slavilor şi latinilor imigraseră recent şi vorbeau prost engleza. Dar, în schimb, invoca în favoarea diferenţei rezultatele la testul *Beta*. Anglofonii obţineau, în medie, un punctaj de 101,6 la *Beta*, iar ceilalţi 77,8. Totuşi, recunoaşte că necunoscătorii de engleză sunt serios penalizaţi la testul *Beta*. O altă corelaţie, care îi putea pune pe gânduri pe ereditarişti, era cea dintre media rezultatelor la teste şi numărul de ani petrecuţi în SUA. De fapt, acest rezultat evidenţia clar faptul că diferenţele înregistrate provin din gradul de adaptare la modul de viaţă american, şi nu din inteligenţa înnăscută. Totuşi, în ciuda acestor evidenţe, cei care au aplicat teste au inventat explicaţii savante pentru a-şi salva prejudecăţile rasiale. În vremea noastră, astfel de concluzii ar fi scandalizat opinia publică, dar în preajma primului război mondial rasismul nu era combătut, ci chiar încurajat de către unele vârfuri ale societăţii americane.

Concluziile analizei rezultatelor la examinări. Foarte îngrijorătoare este afirmaţia după care vârsta mintală medie a examinaţilor este de 13 ani, ştiindu-se că moronii au vârsta cuprinsă între 8 şi 12 ani. Consecinţele şi reacţiile, în urma acestei constatări, nu s-au lăsat mult aşteptate. Preşedintele Colgate-University, G.C. Cutten, spunea în 1922, într-o alocuţiune inaugurală: „Noi nu putem să concepem o formă mai rea a haosului decât o democraţie veritabilă într-o populaţie a cărei vârstă mintală nu depăşeşte 13 ani" (*apud* Gould, 1984, p. 277). Media de 13 ani a avut un impact politic cert. Dar ravagiile sociale ale raportului erau minime, în comparaţie cu cele rasiale şi naţionale. În 1923, C.C. Brigham, asistent la psihologie în cadrul Princeton University, editează o carte, intitulată *A Study of American Intelligence*, care devine instrumentul principal utilizat pentru a traduce în termeni de acţiune socială concluziile testării psihologice din armată. În prefaţa cărţii, semnată de Robert Yerkes, acesta afirmă că nici un cetăţean nu-şi poate permite să ignore pericolul degenerării rasei şi al imigrării, care influenţează negativ progresul şi binele naţiunii americane. Pentru a-şi sprijini tezele rasiste, Brigham face apel la unul dintre cei mai cunoscuţi teoreticieni ai rasismului, americanul Madison Grant, autor al lucrării *The Passing of the Great Race*. Pe baza testelor aplicate în domeniul militar, Brigham ajunge la concluzia că nordicii (locuind în nordul Europei, dominatori, individualişti siguri de ei, rasă de soldaţi, de marinari, de aventurieri, mai ales de religie protestantă) au vârsta mintală de 13,28 ani, alpinii (docili faţă de autoritatea politică şi religioasă, în general de religie romano-catolică) de 11,67 ani, iar mediteraneenii (situaţi în sudul Europei) de 11,43 ani. Influenţa mediului este acceptată doar pentru a discrimina

rasa nordică în englezi şi non-englezi. Alţi propagandişti ereditarişti s-au servit de rezultatele testelor aplicate în domeniul militar pentru a justifica segregaţia rasială şi limitarea accesului negrilor la învăţământul de înalt nivel. La şapte ani după declaraţiile categorice în favoarea rasismului, Brigham îşi schimbă complet viziunea şi recunoaşte că un rezultat al testului nu putea fi reificat într-o entitate localizată în craniul omenesc. El afirmă că cele două teste, *Alpha* şi *Beta*, măsoară lucruri distincte şi nici unul dintre ele nu are coerenţă internă. În al doilea rând, Brigham recunoaşte că testele măsoară gradul de adaptare la limba şi la cultura americană, şi nu inteligenţa înnăscută. Rareori, în literatura de specialitate psihologică, s-au văzut astfel de retractări categorice.

Referiri la deteriorarea inteligenţei naţionale. Folosirea testelor *Army Alpha* şi *Army Beta* a reaprins discuţia despre deteriorarea nivelului de inteligenţă naţional. Despre jumătate dintre bărbaţii albi testaţi în armată se spune că aveau inteligenţa egală cu aceea a copiilor de 13 ani şi chiar mai scăzută, situaţia fiind şi mai rea pentru soldaţii negri. Răspunsul lui Goddard la această situaţie era că populaţia cu abilităţi mintale scăzute nu ar trebui să voteze. Goddard, Terman şi Yerkes erau foarte preocupaţi de deteriorarea inteligenţei naţionale, văzând drept cauze imigrarea şi faptul că indivizii inferiori din punct de vedere intelectual se reproduceau ca indivizii normali. Ca un lucru comun acelor timpuri, Yerkes (1923) credea că numeroase probleme ale naţiunii erau cauzate de persoanele cu inteligenţă scăzută şi că politica de imigrare nu face decât să agraveze acest lucru.

Din fericire, poziţia nativist-extremistă pe care Goddard, Yerkes şi Terman o reprezentau nu a rămas neschimbată. Din ce în ce mai mulţi au realizat că performanţa la *testele de inteligenţă* poate fi explicată prin factori precum experienţa timpurie şi educaţia. Mai curând decât simpla măsurare a inteligenţei native, pe care o estimau aparent, testele măsurau achiziţiile personale şi influenţa circumstanţelor vieţii. Rezulta că o persoană era privilegiată mai mult în termeni de îmbogăţire a experienţei şi educaţiei pe care se bazează un scor înalt la testele de inteligenţă. Într-o recentă lucrare de specialitate, *The Bell Curve : Intelligence and Class Structure in American Life* (1994), Richard Herrnstein şi Charles Murray ilustrează câteva dintre credinţele timpurii despre inteligenţă, acceptate de Galton, Goddard, Cattell, Terman şi Yerkes. Herrnstein şi Murray au structurat lucrarea lor în jurul unor concluzii despre inteligenţă care erau „dincolo de dispute". Erau menţionate şase puncte:

1. Oamenii diferă în privinţa factorului general de abilitate cognitivă.
2. Toate testele standardizate de aptitudini sau de achiziţii academice măsoară respectivul factor general pe câteva grade, dar testele IQ, expres desemnate în acest scop, măsoară cu mai multă acurateţe.
3. Scorurile IQ se potrivesc, într-o primă instanţă, cu ceea ce oamenii înţeleg atunci când folosesc cuvântul „inteligent" sau „deştept" într-un limbaj obişnuit.
4. Scorurile IQ sunt stabile, deşi nu perfect, pentru o durată mai mare decât viaţa medie a unei persoane.
5. Administrate cum se cuvine, rezultatele testelor IQ nu sunt deplasate sistematic în grupurile sociale, economice, etnice şi rasiale.
6. Abilitatea cognitivă este substanţial ereditară, aparent nu mai puţin de 40 de procente şi nu mai mult de 80 de procente (*apud* Huteau şi Lautrey, 1995, p. 64).

Autorii amintiţi anterior susţin că există o dispută în Statele Unite ale Americii în privinţa ideii potrivit căreia cele mai bune slujbe şi cele mai mari câştiguri le-ar reveni celor mai inteligenţi indivizi, şi anume elitei cognitive. Cei mai slab dotaţi intelectual

sunt condamnaţi la munci de slugă într-o economie bazată pe informaţie, dacă reuşesc şi să găsească o astfel de slujbă. Acest fapt, asociat concepţiei că inteligenţa este în mare măsură moştenită, duce la concluzia după care America are o problemă majoră, constând în structura de clasă economică bazată pe inteligenţa moştenită. Autorii nu oferă o soluţie la problema structurii de clasă bazată pe inteligenţa moştenită, dar savanţii de la începutul secolului XX, precum Galton, Cattell, Goddard, Terman şi Yerkes, au descris o problemă similară şi au sugerat că soluţia ar reprezenta-o descurajarea reproducerii indivizilor cu inteligenţă scăzută. De fapt, respectivele puncte de vedere despre inteligenţă au fost şi au rămas subiecte fierbinţi de dezbatere.

Apariţia profilului psihologic. În Europa, un pas important în direcţia sistematizării datelor pe care le poate obţine un psiholog în cadrul unui examen l-a făcut psihologul rus Rossolimo, prin crearea profilului psihologic. El proiectează un tip de examen în care rezultatele sunt reprezentate printr-un grafic numit *profil psihologic*. Acest profil, în concepţia originală a autorului, cuprinde o listă de 150 de probe repartizate în 15 serii, fiecare probă fiind în 10 grade, clasate de la 1 la 10, după dificultatea crescătoare a rezultatelor obţinute la copiii normali de diferite vârste. Rossolimo studiază trei tipuri de funcţii psihice: de execuţie, de elaborare şi de achiziţie. Acest prim profil psihologic este încă empiric şi discutabil. W. Stern aduce o îmbunătăţire consistentă, substituind notei rangul obţinut de subiect, în ansamblul subiecţilor de aceeaşi categorie. De asemenea, pe parcurs, numărul gradaţiilor a variat, de cele mai multe ori, între 5 şi 10. Prezentăm un exemplu de profil psihologic preluat după S. Chamboulant (1951, p. 136).

Probe	Valori	Decile									
		1	2	3	4	5	6	7	8	9	10
Inteligenţa											
Logică											
Numerică											
Raţionament											
Judecată											
Memoria											
Cifre											
Cuvinte cuplate											
±											
Recunoaşteri											
Povestiri											
Semne											
Consemne											
Atenţie											
Concentrată											
Distributivă											
Comportament											

Figura 2. *Profil psihologic de tip Rossolimo*

4.2. Dezvoltarea mişcării testelor până la al doilea război mondial

Plecând de la opera lui Binet şi de la marele eveniment istoric, primul război mondial, mişcarea testelor va cunoaşte în perioada interbelică o dezvoltare considerabilă. Testele vor fi aplicate la populaţii din ce în ce mai diverse şi din ce în ce mai numeroase nu doar în SUA, ci şi în toate ţările industrializate dezvoltate (deşi într-o mai mică măsură). Se poate vorbi totuşi despre o *americanizare* a cercetării ştiinţifice în domeniul creării şi dezvoltării acestora, nu în sensul de activitate exclusiv americană, ci de activitate predominant americană. Experienţa testării psihologice în primul război mondial, de amploare neegalată până astăzi, constituie un moment de referinţă în istoria psihodiagnosticului. Iniţiativa preluată de americani în privinţa construcţiei şi perfecţionării probelor psihologice nu a fost cedată până astăzi, decât cu foarte puţine excepţii.

Tehnicile de construcţie a testelor vor cunoaşte un progres sensibil, dar bazele lor teoretice vor rămâne incerte. Această dezvoltare inegală a componentei teoretice a testului psihologic va sta la originea unor vii dezbateri ştiinţifico-ideologice.

Dezvoltarea testelor în America este deosebit de accentuată în sistemul educativ, fiind sistematic utilizate pentru detectarea debililor mintal. Frecvent, se aplică în selecţia elevilor supradotaţi (definiţi simetric faţă de debili). Practica utilizării testelor, pentru a constitui clase omogene, se generalizează (puternica omogenitate era atunci considerată un factor favorabil al învăţării). Folosirea testelor de inteligenţă, pentru selecţia şi orientarea elevilor din licee, devine un loc comun. Testele îşi găsesc utilizare şi în învăţământul superior, în anul 1926, o dată cu primele versiuni ale *Scholastic Aptitude Test* (*SAT*), construit de către C.C. Brigham la cererea adresată de College Entrance Examination Board (CEEB), pentru a decide admiterea studenţilor la universitate. *SAT*, constant remaniat, este utilizat permanent şi masiv la ora actuală (el comportă o secţiune verbală – comprehensiunea testelor, antonime, analogii verbale, completare de fraze – şi o secţiune *matematică*, presupunând cunoştinţele matematice, dar nu specializate). Pentru a gira aplicaţiile şi adaptările acestui test, a fost constituit la Princeton un organism special, care va deveni în 1947 Educational Testing Service (ETS), considerat astăzi cea mai reputată instituţie în materie de cercetări şi aplicaţii pentru problema testului şi evaluării şcolare. Numărul de elevi şi studenţi testaţi prin acest sistem este de ordinul milioanelor (testul creat de Robert Yerkes, intitulat *National Intelligence Test*, care nu reprezintă decât un test de inteligenţă printre altele, este aplicat, în 1920, la 7 milioane de elevi). Aplicaţiile testelor în alte domenii sunt, de asemenea, numeroase. Atât pentru selecţie, cât şi pentru orientarea profesională se utilizează teste de inteligenţă. Inteligenţa a fost şi este considerată o „regină" a aptitudinilor umane, iar perfecţionarea testului a pornit, în cea mai mare parte, de la preocupările vizând diagnosticarea acesteia. Se construiesc teste de aptitudini variate, care pot să evalueze componentele inteligenţei sau ale aptitudinilor particulare cerute în exercitarea diferitelor profesiuni. Practica diagnosticului prin intermediul testelor individuale se dezvoltă la fel de mult în sectorul sănătăţii (spitale de psihiatrie, clinici de consultaţii pentru copii), precum şi în sectorul judiciar. Referitor la psihologia militară, ea cotinuă să fie practicată şi după încheierea primului război mondial (aproximativ 2.000 de persoane activau, în anul 1918, în serviciul psihotehnic al armatei americane).

Chiar dacă dezvoltarea prioritară se realizează în Statele Unite ale Americii, se poate considera că, între cele două războaie mondiale, mişcarea testelor psihologice a devenit internaţională. Progresele europenilor sunt şi ele notabile. În 1920, este fondată, la iniţiativa lui Claparède, Asociaţia Internaţională de Psihotehnică. Aceasta va organiza un prim congres în acelaşi an, la Geneva, unde au fost reprezentate şapte ţări. Numeroase laboratoare şi institute consacră esenţialul activităţii lor creării de teste mintale, mărturisind, în acest fel, vigoarea mişcării testelor, chiar dacă volumul aplicaţiilor efective rămâne încă modest. În Franţa, se crează Institutul de Psihologie al Universităţii din Paris (în anul 1920), un laborator de psihologie industrială în cadrul spitalului Henry-Rousselle din Paris (în 1923), un laborator psihotehnic la Societatea de transport în comun din regiunea pariziană (în 1924), un laborator de psihologie aplicată la École Pratique des Hautes Études (în 1927). Prima instituţie de editare a testelor din Franţa, Établissements d'applications psychotechniques (EAP), se înfiinţează în anul 1927, iar un an mai târziu, Institutul Naţional de Orientare Profesională (INOP). În 1932, căile ferate din nord îşi înfiinţează servicii psihotehnice şi, după un an, cele din est le urmează exemplul.

Dezvoltarea testării psihologice în Rusia sovietică merită o atenţie particulară. După lovitura de stat din octombrie 1917, autorităţile îi acordă psihotehnicii o importanţă deosebită, fiind considerată o tehnică ştiinţifică bine fundamentată şi susceptibilă de a ajuta la prosperitatea ţării. Dezvoltarea sa este activ susţinută de către factorii de conducere. În anul 1928, autorităţile sovietice afirmă că psihotehnica trebuie să contribuie la realizarea obiectivelor de dezvoltare ale Uniunii Sovietice. În jurul anului 1930, exista deja în URSS o reţea de centre de cercetare asupra testelor şi se editau trei reviste de specialitate psihotehnică. La al VII-lea Congres Internaţional de Psihotehnică, desfăşurat la Moscova în 1931, participă 200 de reprezentanţi sovietici, iar psihologii din Occident sunt impresionaţi de mijloacele de care dispun colegii lor şi de calitatea realizărilor lor. Testele sunt aplicate frecvent în uzine şi în şcoli.

4.2.1. *Dezvoltările conceptuale şi metodologice*

Construcţia testelor de inteligenţă pune probleme conceptuale (ce este inteligenţa?) şi probleme metodologice (cum o vom măsura?). Pentru psihologii europeni de la începutul secolului XX, Binet şi Spearman, aceste două probleme nu sunt diferenţiate, iar analiza funcţionării intelectuale şi măsura eficienţei sale sunt tratate în ansamblu. Cu câteva excepţii notabile, se va abandona, practic, studiul funcţionării inteligenţei sau se va desfăşura independent de măsurarea inteligenţei. Totuşi, în această perioadă apare ideea conform căreia este greşit ca inteligenţa să fie considerată o dimensiune unică. În acelaşi timp, metodologia de construcţie şi analiză a testelor cunoaşte dezvoltări importante de care vor beneficia şi alte discipline. Această situaţie, de disociere accentuată, este explicată prin întâlnirea unor factori teoretici şi a unor factori sociali. Totuşi, din 1930, testele încep să fie fundamentate pe o mai bună înţelegere a metodelor de eşantionare şi a psihometriei, în general.

Pe plan teoretic, o mare dezvoltare cunoaşte behaviorismul, situaţie care deturnează psihologii de la analiza funcţionării intelectuale. Behaviorismul se consacră studiului legilor elementare ale învăţării (faimoasele relaţii stimul – răspuns), procedurilor de întărire, dar aproape deloc conduitelor complexe care sunt presupuse de termenul de *inteligenţă*.

Pe plan social, marea cerere de teste determină privilegierea construcţiei instrumentelor care „merg" sau „par să meargă", adică se dă o mai mare importanţă eficacităţii practice, decât cunoaşterii mecanismelor teoretice. Aplicaţiile propuse prin behaviorism, care vizează modificarea comportamentului, se opun testelor şi nu le cuprind pe acestea.

4.2.2. *Concepţia unidimensională şi pluridimensională a inteligenţei*

Psihologii care au construit testele armatei americane considerau inteligenţa un ansamblu de funcţiuni sau de capacităţi pe care testele trebuie să le asimileze. În acest sens, ei se situează aproape în descendenţa lui Alfred Binet, care nu a pretins că măsoară o „inteligenţă generală". Inteligenţa este definită, astfel, prin capacitatea de a înţelege limbajul, de a raţiona asupra relaţiilor cantitative şi semantice, de a emite judecăţi practice, de a infera reguli şi regularităţi, de a memoriza. Rămâne de explicat caracterul unitar al inteligenţei:

- pentru ce şi cum sunt aceste capacităţi puternic integrate?
- pentru ce şi cum tind capacităţile să se manifeste în gradiente apropiate la aceiaşi subiecţi?

E.L. Thorndike este unul dintre cei care au furnizat răspunsurile cele mai elaborate la aceste chestiuni. El explică faptul că inteligenţa desemnează, mai întâi, capacitatea de a stabili conexiuni între idei. Or, pentru a răspunde la sarcinile complexe ale testelor de inteligenţă, este necesar să se stabilească astfel de conexiuni. Manifestările de inteligenţă depind, aşadar, de stocul de conexiuni disponibile. În anul 1982, J.B. Carroll face precizarea că această explicaţie este asemănătoare cu aceea a lui Thomson, un psiholog scoţian care, în anii '20, motivează corelaţiile între teste prin identitatea proceselor solicitate de ele. Fiecare sarcină, pentru a fi rezolvată, presupune activarea unui ansamblu de procese, iar corelaţiile între sarcini sunt cu atât mai ridicate cu cât legăturile între ansamblurile de procese sunt mai extinse. Concepţia este originală, în măsura în care ea se fondează pe unicitatea inteligenţei, şi nu pe pluralitatea proceselor intelectuale.

Fără îndoială, englezul Charles Spearman (1863-1945) este cel care va justifica cel mai bine ideea de inteligenţă generală. Teoria lui Spearman este în strânsă legătură cu o tehnică de analiză a matricelor de corelaţii, analiza factorială, ale cărei baze le pune în anul 1904 şi care cunoaşte un mare succes şi în afara psihologiei.

Logica intervenţiei acestor tehnici cantitative rezidă în reducerea numărului total de variabile ce pot interveni în explicaţia rezultatelor cercetărilor psihologice, număr care la un moment dat părea să fie limitat doar de capacitatea umană de a da nume pentru diverse calităţi psihice supuse analizei. În situaţia în care cercetătorii nu se pot pune de acord asupra variabilelor fundamentale, speranţa unui progres ştiinţific este foarte redusă. „Dacă fiecare psiholog alege şi defineşte propriile sale variabile fără a se interesa cum aleg şi definesc alţi psihologi variabilele lor, atunci rezultatul de neevitat va fi un enorm haos de variabile neordonate şi nestructurate, care lasă să apară comportamentul omenesc confuz, în loc ca el să apară clar şi raţional", spune unul dintre teoreticienii de marcă ai testului psihologic[1].

1. Vezi Paul Horst, *Messung und Vorhersage*, Verlag Julius Beltz, Weinheim, Berlin, Basel, 1971, p. 170.

Analiza factorială nu este o tehnică unică, ci un ansamblu cu mai multe faţete, dintre care enumerăm următoarele:

- analiza în componente principale;
- analiza în factori comuni şi unici;
- analiza factorială a imaginilor a lui Guttman;
- analiza factorială alfa a lui Kaiser şi Caffrey;
- analiza factorială a corespondenţelor a lui J.P. Benzécri.

Analiza factorială constă, în esenţă, în explicarea corelaţiilor între variabilele observate prin proximitatea acestor variabile observate faţă de anumite variabile ipotetice (sau prin „saturarea" variabilelor observate prin variabilele ipotetice). Se considera că aceşti factori reprezintă o realitate subiacentă a conduitei. Spearman, ale cărui lucrări fundamentale apar în anul 1923 (*The Nature of „Intelligence" and the Principles of Cognition*) şi în anul 1927 (*The Abilities of Men*), consideră că toate sarcinile intelectuale sunt mai mult sau mai puţin saturate în acelaşi factor, factorul general sau *g*, fiecare fiind saturată şi în factorul care îi este propriu, factorul specific *s* (din acest motiv, teoria lui Spearman a fost calificată drept bifactorială). El caută să precizeze natura factorului *g*. Acest factor general va corespunde capacităţii de a stabili şi de a aplica relaţii şi este determinat de cantitatea de energie nervoasă de care dispune subiectul.

Rafinarea concepţiei analizei factoriale se face spre anii '30 (Spearman, 1927; Kelly, 1928; Thurstone, 1938).

În anii '30, Louis Leon Thurstone (1887-1955) va dezvolta, în SUA, tehnici de analiză factorială şi va preciza fundamentele lor matematice (algebra matriceală). Cercetările sale vor determina luarea în considerare a concepţiilor unitare ale inteligenţei, pe atunci dominante, indiferent dacă se calculează un IQ sau se evaluează factorul *g*. Măsurarea inteligenţei este considerată a fi reprezentată mai mult decât de un singur scor. Chiar din 1924, Thurstone a prezentat un test în care se calculează două scoruri: un scor *lingvistic*, scorul L, şi un scor *cantitativ*, scorul Q. Această nouă optică de scorare este adoptată de American Council on Education Psychological Examination. În 1927, *Scholastic Aptitude Test* permite calcularea unui scor *verbal* şi a unui scor *matematic*, primul corelând înalt cu gradul de avansare la cursurile de limba engleză şi istorie, al doilea cu cel de la matematică şi desen tehnic.

Analiza factorială va pune în evidenţă distincţii pertinente. L.L. Thurstone arată că teoria bifactorială a lui Spearman nu este validă decât pentru eşantioane de sarcini particulare. Dacă o persoană se interesează de eşantioane de conduite intelectuale nedeformate de către influenţele sistematice (deformări desemnate, în ultima vreme, prin termenul *biais*), atunci ea trebuie să explice corelaţiile observate, făcând apel la mai mulţi factori de grup, adică la factori care saturează puternic mai multe variabile, fără a fi generali (teoria multifactorială). L.L. Thurstone prezintă ansamblul modelului său în 1935, în *The Vector of the Mind*. El stabileşte o listă de nouă factori sau capacităţi care caracterizează formele de inteligenţă: P – viteză perceptivă; N – capacitate numerică; V – comprehensiune verbală; W – fluiditate verbală; S – capacitate de vizualizare spaţială; M – capacitate de memorizare; I – inducţie; D – deducţie; R – raţionament. Ulterior, Thurstone modifică această listă.

În final, cercetătorul american explică destul de puţin funcţionarea mintală propriu-zisă, dar arată că, nici în situaţia în care se limitează situaţia de observaţie la nişte teste creion-hârtie, inteligenţa nu este un concept unitar. Consecinţele pentru evaluare sunt

importante: subiectul nu mai trebuie să fie caracterizat printr-un scor pe o dimensiune unică, ci printr-un profil definit prin scoruri la mai multe dimensiuni. Am amintit anterior că psihologul rus Rossolimo a avut primul ideea, în anul 1912, de a prezenta rezultatele unui subiect sub forma unui profil normat.

Reprezentanţii şcolii psihometrice engleze, mai ales C. Burt şi P.E. Vernon, arată că, plecând de la corelaţii între factori primari, se poate reintroduce noţiunea de *factor general* şi propun un model ierarhic al organizării factorilor. Influenţa acestui factor general şi interpretarea sa sunt încă, la ora actuală, obiecte de discuţii ştiinţifice. Factorul *g* este întotdeauna relativ la un eşantion de sarcini (el nu este consecinţa covariaţiei pozitive între respectivele sarcini). Dacă existenţa sa este bine stabilită când sarcinile fac apel la achiziţii verbale şi la diverse forme de raţionament inductiv sau deductiv (inteligenţă abstractă ori verbală), ea este, în schimb, mult mai slab determinată dacă se introduc eşantioane de sarcini ale probelor în raport cu inteligenţa practică sau inteligenţa socială. De asemenea, rămânând în domeniul inteligenţei abstracte sau verbale, putem cerceta ponderea acestor factori în variabilitatea performanţelor, apreciindu-i ca fiind mai mult sau mai puţin importanţi.

Majoritatea psihologilor din această perioadă sunt convinşi că inteligenţa este ereditară şi unii dintre ei văd în genetica mendeliană baza teoretică cea mai sigură a psihologiei diferenţiale a inteligenţei. Ei diferă totuşi prin importanţa pe care o acordă factorilor neereditari. Pentru unii, minoritari (este deja cazul lui Binet), inteligenţa nu este independentă de condiţiile sociale de existenţă şi de educaţia familială şi şcolară, fiind, în consecinţă, educabilă.

4.2.3. *Teoria măsurii şi construcţia testelor*

Dacă cercetările teoretice se dezvoltă relativ puţin, în schimb sunt elaborate numeroase lucrări metodologice. Ele sunt realizate mai ales în domeniul teoriei măsurării şi în domeniul principiilor de construcţie a testelor, apelând foarte mult la statistica descriptivă. Problemele puse de construcţia testelor de inteligenţă vor fi, de altfel, una dintre importantele surse ale dezvoltării statistice descriptive.

Corelaţiile se păstrează în prim-planul operaţiilor statistice. Chiar din anii '20, tehnicile corelaţionale sunt frecvent utilizate. Yerkes, în raportul său din anul 1921 asupra testării militarilor americani din timpul primului război mondial, prezintă una dintre primele aplicaţii ale tehnicii de regresie multiplă.

Pentru a măsura inteligenţa, trebuie să se justifice existenţa continuităţii, permiţând cel puţin posibilitatea de a ordona indivizii. Binet utiliza scala de vârste pentru a fonda un astfel de continuum. Dar el s-a mulţumit să verifice la modul superficial vârsta medie a achiziţiilor noţiunilor utilizate în test (Terman va proceda la fel) fără a se preocupa de ordinea achiziţiilor care, pentru ca măsura să fie fondată, trebuie să fie constantă. Vârsta mintală are un sens dacă o aceeaşi vârstă mintală este obţinută, aproape în acelaşi mod, de către toţi cercetătorii. Analizele sistematice ale *pattern*-urilor de răspuns, în scopul verificării coerenţei lor, vor începe să fie efectuate la începutul anilor '30 şi va trebui să vină anul 1941 pentru ca Guttman să-şi prezinte metoda de *analiză ierarhică*. În cazul în care testele sunt constituite din itemi greşiţi sau reuşiţi, ordonarea subiecţilor poate fi fondată pe următorul principiu: pentru ca doi subiecţi să aibă acelaşi nivel, nu este suficient ca ei să aibă acelaşi număr de reuşite; mai trebuie ca reuşitele să se refere la aceiaşi itemi. Crearea acestui tip de test, după primul război mondial, presupune existenţa

unei dimensiuni unice care se stabileşte plecând de la corelaţiile interitemi: dacă într-un ansamblu de itemi există o corelaţie notabilă între itemi, se poate considera că itemii măsoară o aceeaşi dimensiune. Acest mod de raţionament va fi sistematizat prin utilizarea tehnicilor de analiză factorială. Începând din anii '30, metodologia de construcţie a testelor atinge un înalt grad de complexitate.

În anul 1982, J.B. Carroll rezumă situaţia din domeniul construcţiei testelor, aşa cum se prezenta ea în 1935 (anul în care s-a înfiinţat Psychometric Society – cu revista sa, intitulată *Psychometrica* – regrupând specialiştii în metodologia testelor). Autorul amintit susţine că „progrese tehnologice au permis producţia unui nou tip de instrument social care, în climatul epocii, a fost larg acceptat; totuşi, natura exactă a acestui instrument şi semnificaţia sa veritabilă rămân încă neexplorate în mare măsură" (Carroll, 1993, p. 57).

4.2.4. *Diversificarea testelor de inteligenţă şi generalizarea psihometriei*

Testele de inteligenţă se vor diversifica pornind de la prototipul testului elaborat de Binet şi Simon, de la adaptarea sa de către Lewis Terman, sub presiunea nevoilor practicii şi prin asimilarea reflecţiilor teoretice asupra inteligenţei.

Totuşi, modelul Binet-Simon al examinării inteligenţei nu a fost preluat doar în America, ci şi în numeroase ţări din Europa. Preluările sunt atât de numeroase încât stabilirea bibliografiei acestei teme a constituit o direcţie de preocupare pentru mari psihologi, cum ar fi O. Decroly şi J. Degand, O. Bobertag, H. Goddard, L.M. Terman, W. Stern, E. Meumann, F. Kuhlmann, E.A. Kirkpatrick[1]. Majoritatea celor preocupaţi au acceptat sistemul lui Binet, dar o mică parte au început să-i facă reproşuri imediat ce au luat cunoştinţă de el (de exemplu E. Meumann). Criticile au fost direcţionate mai ales asupra ordonării sarcinilor pe trepte de vârstă (cercetătorii ce favorizau în concepţie diagnosticul bine precizat). O mare parte a dificultăţilor reieşeau din problemele tehnice confruntate cu o experienţă particulară. Unii autori traduceau pur şi simplu testul din franceză, alţii operau schimbări doar atunci când conţinutul sarcinii nu era foarte clar, datorită particularităţilor situaţionale franceze, doar o mică parte dintre autori permiţându-şi intervenţii profunde şi inventând noi sarcini. Dificultatea principală consta, în acest caz, în stabilirea limitelor de vârstă pentru anumite sarcini.

O parte dintre cei care au preluat sistemul examinau, pe cât posibil exact, copii de 8, 9, 10 ani, alţii de 8-9 ani şi alţii de 9-10 ani, neclarităţile rezultând din metodologia de selecţie a subiecţilor destinaţi examinării. Unii examinatori aplicau testul doar la copiii denumiţi de profesori „normali" sau „satisfăcători", în timp ce alţii cereau clase întregi, şcoli întregi sau chiar copiii dintr-un oraş întreg. Nu se lua în considerare originea socială a copiilor, în majoritatea examinărilor.

Problemele statistice nu erau în nici un caz clare, mai ales cele care se refereau la cotele procentuale de copii ce trebuiau să rezolve o anumită sarcină (pentru ca această sarcină să fie considerată specifică unei vârste). O. Bobertag, J.E. Wallin, L.M. Terman şi H.G. Childs au propus să fie luate în considerare cote procentuale variind între 75% şi 100%.

1. O listă completă se poate găsi în Carl Josef Groffmann, „Die Entwicklung der Inteligenzmessung", în *Handbuch der Psychologie in 12 Bände, 6 Band, Psychologische Diagnostic*, Verlag der Psychologie, Dr. C. Hogrefe, Göttingen, 1964, p. 169.

În ciuda unor reproşuri la adresa creaţiei lui Binet şi Simon, şi-a făcut loc totuşi ideea că începutul hotărâtor era făcut şi domina optimismul în privinţa posibilităţii realizării unei metode de psihodiagnostic cert folositoare, cu câteva probe adecvate în privinţa solicitării.

În Germania, este de remarcat apariţia unor concluzii sistematice în privinţa operei lui Binet, datorate lui W. Stern, coordonatorul cercetărilor de psihologie diferenţială, care a realizat la începutul secolului XX (în aprilie 1911), pentru Congresul psihologilor din Berlin, un referat de sinteză despre examinarea inteligenţei (publicat în acelaşi an). Cu această ocazie, apare celebra definiţie: „Inteligenţa este aptitudinea generală a unui individ de a-şi regla conştient gândirea la noi solicitări; ea este o aptitudine psihică generală de adaptare la noi sarcini şi condiţii de viaţă"[1].

Cunoscuta formulă a coeficientului de inteligenţă apare în această lucrare. Fundamentarea IQ-ului se bazează pe cercetările cercului ştiinţific al lui W. Stern din Breslau, din care făceau parte O. Bobertag, autorul celei mai serioase revizii naţionale (germane) a scalei Binet-Simon, şi psihiatrii F. Chotzen şi F. Krammer, care au efectuat cercetări pe debili mintal. Alfred Binet era convins că mai mult de doi ani diferenţă între vârsta cronologică şi vârsta mintală era un semn al debilităţii mintale, dar din cercetările lui Krammer şi Chotzen rezultă că, pentru diagnosticarea debilităţii, diferenţa necesară între vârsta mintală şi cea cronologică nu rămâne constantă, ci se măreşte o dată cu vârsta, astfel încât compararea debilităţii mintale de diverse vârste prin intermediul standardului de inteligenţă utilizat până atunci, exprimat în luni, nu este posibilă. Stern a putut să indice, pe baza recalculărilor efectuate pe rezultatele lui Chotzen, că valoarea IQ-ului prezintă o măsură foarte realistă a retardării în comparaţie cu cea folosită până atunci. El recomandă folosirea coeficientului de inteligenţă doar atunci când sunt necesare comparaţii între indivizi normali şi anormali (având inteligenţa copiilor) pentru diferite vârste.

În privinţa obiecţiilor la adresa testului de inteligenţă, menţionăm că o parte dintre ele au fost ridicate deja în perioada vieţii lui Binet, unele neputând să fie lămurite nici până astăzi (între acestea intră condiţiile extraintelectuale ale randamentului la test, care au fost tratate incomplet). Unele obiecţii au fost ridicate între 1920 şi 1930, cele mai multe argumentări prezentând şi propuneri concrete de îmbunătăţire, care nu au putut fi realizate imediat.

Încet, dar sigur, centrul preocupărilor privind perfecţionarea testului psihologic se mută în SUA. Americanii excelează prin mulţimea lucrărilor consacrate testului, prin înclinaţia lor spre statistică, prin numărul uriaş de subiecţi examinaţi şi prin faptul că rezultatele erau prelucrate nu numai pentru procedeul de examinare în ansamblu, ci şi pentru trepte individuale de vârstă. Aici este locul în care se creează şi se probează noi sarcini pentru grupe de vârstă superioare.

O dată cu testele psihologice create pentru examinarea recruţilor din SUA, în anul 1917 apare o nouă formă de test, şi anume testul colectiv de inteligenţă generală, căruia i se spune în mod obişnuit *test creion-hârtie*. H. Ebbinghaus avea idee de aceste teste, dar primele realizări sunt *Army Alpha* şi *Army Beta*. Testelor colective li se datorează dezvoltarea masivă a practicării testelor. Acestea diferă de scala Binet-Simon şi de scala Stanford-Binet a lui Terman, în privinţa facilităţii lor de aplicare. Pe când itemii celor două tipuri de teste amintite anterior sunt tipici pentru o anumită vârstă, itemii testelor

1. *Ibidem*.

colective nu se referă la o anumită vârstă. Ele sunt simple, au dificultate progresivă şi scorul obţinut este un număr de puncte. Măsura eficienţei este diferenţiată de aceea a dezvoltării. În majoritatea itemilor testelor elaborate de Binet-Simon sau Terman, subiectul furnizează liber un răspuns cotat de către examinator (în funcţie de nivelul de dezvoltare pe care el îl revelează). Micile probleme care constituie testele colective pot fi reuşite sau eşuate şi, în general, sarcina subiectului se limitează la a alege un răspuns dintre mai multe care îi sunt propuse. Aceste diferenţe nu trebuie totuşi să mascheze o intenţie clară, şi anume aceea de a situa subiectul pe un continuum unic.

Lucrările realizate prin intermediul tehnicilor de analiză factorială dau naştere la o categorie particulară de teste colective, *testele factoriale*. În cadrul teoriei lui Spearman, s-au construit teste de factor *g*. Dacă aceste teste au drept obiectiv măsurarea inteligenţei generale, conţinutul lor este foarte liber în raport cu al celor care le precedă. Acest conţinut nu rezultă dintr-o eşantionare imprecisă a diverselor capacităţi, ci este strict determinat de interpretarea pe care Spearman o dă factorului *g*. Toate testele de factor *g* sunt constituite din itemi în care se pune problema descoperirii legilor de transformare (raţionament inductiv) dintr-o serie de desene geometrice aparent fără semnificaţie. Cel mai cunoscut test de factor *g* este *matricea progresivă Raven*, care apare într-o primă ediţie în anul 1938. De asemenea, R.B. Cattell construieşte, în 1940, un test de factor *g*. Prin prisma teoriei lui L.L. Thurstone, s-au construit teste destinate evaluării eficienţei în diverse sectoare ale funcţionării intelectuale. În 1938, apare prima versiune (o versiune experimentală) a *Primary Mental Abilities Tests*, aparţinând lui L.L. Thurstone. Testul caracterizează subiectul după şapte scoruri.

În acelaşi timp cu testele creion-hârtie, cu cele verbale sau nonverbale, apar testele numite de *performanţă*. De această dată, subiecţii trebuie să modifice un material concret care le este prezentat, în scopul de a atinge un obiectiv fixat (de exemplu, a reproduce un desen prin aranjarea cuburilor, ca în puzzle). După primele tentative ale lui Knox (care cerceta un mijloc pentru a testa imigranţii neanglofoni din SUA), prima scală de performanţă a fost prezentată de Pintner şi Paterson în 1917. Testul cuburilor lui Kohs datează din 1923, scala lui Grace Arthur din 1930, iar aceea a lui Alexander din 1935. Anumite scale de performanţă permit evaluarea inteligenţei analfabeţilor sau a celor care nu înţelegeau engleza. Ele nu erau pe atunci decât simple substitute ale probelor de inteligenţă generală. Pentru unii cercetători, aceste probe, ca şi testele colective de inteligenţă generală nonverbale sau testele de factor *g*, permit măsurarea inteligenţei independent de nivelul cultural (se observă rapid că acest lucru nu era întru totul posibil). Pentru alţii, cele mai realiste scale de performanţă evaluează câteva aspecte ale unei forme particulare de inteligenţă, numită *inteligenţă practică*.

Unii autori afirmă că în jurul anului 1940 se poate vorbi despre o nouă treaptă a dezvoltării testului de inteligenţă (vezi Huteau şi Lautrey, 1995). Alţi autori vorbesc de anul 1930, ca moment de cotitură în istoria testelor. Considerăm că, în general, pe parcursul deceniului patru s-au produs înnoiri semnificative în domeniul construcţiei testelor psihologice. Ansamblul procesului de examinare psihologică prin testul psihologic devine mai dinamic, iar interacţiunea părţilor mai elastică.

Nevoile activităţii de psihodiagnostic clinic au fost la originea invenţiei unei serii de teste individuale, de exemplu, testul de structurare vizuală a lui Bender (1938), în care subiectul trebuie să recopieze figurile geometrice, sau testul de clasament al

obiectelor al lui Gelb şi Goldstein (1925), în care trebuie să se ordoneze obiecte diferite prin culoarea, forma şi materia lor.

Caracteristicile de dinamism şi elasticitate ale testului modern se fac din plin simţite la apariţia testului de inteligenţă creat de David Wechsler. Testul este „pornit" în anul 1932, dar anul apariţiei este 1939, când este dezvoltat la Bellevue Hospital din New York, cu mari speranţe în privinţa contribuţiei la diagnosticul psihiatric. Considerăm acest test ca aparţinând perioadei dintre cele două războaie mondiale, deoarece construcţia sa a început în acea vreme, perfecţionările continuând zeci de ani după cel de-al doilea război mondial. Testul Wechsler-Bellevue este compus din şase subteste verbale şi cinci subteste de performanţă. În 1946, apare Wechsler-Bellevue II, construit ca un echivalent şi o alternativă la varianta anterioară. Această formă nu a fost niciodată standardizată în mod adecvat. În 1949 apare Wechsler Intelligence Scale for Children, sau pe scurt WISC, care a fost o extensie a testului Wechsler Bellevue II pentru copiii de la 5 la 15 ani şi 11 luni. Ca o carenţă a acestei forme, putem considera faptul că eşantioanele de standardizare reunesc numai albi. În 1955, apare WAIS (Wechsler Adult Intelligence Scale), care este o restandardizare a lui Wechsler-Bellevue cu o serie de itemi revizuiţi pentru adulţi cu vârste mai mari de 16 ani. În 1967, apare WPPSC (Wechsler Preschool and Primary Scale for Children), destinat copiilor de la 4 la 6 ani şi jumătate. A fost primul test de adecvare a eşantionului minorităţilor rasiale. În 1974, apare WISC-R (Wechsler Intelligence Scale for Children Revised) cu itemi transformaţi în informaţie, vocabular şi completarea de desene. Noile norme includ afro-americani şi alte minorităţi, în proporţii adaptate. Este destinat copiilor de la 6 ani la 16 ani şi 11 luni. În 1981, apare WAIS-R, o revizie a WAIS-ului cu 80% din itemi reţinuţi şi norme aduse la zi. În 1989, apare WPPSC-R, o revizie a WPPSC cu itemii inadecvaţi eliminaţi, destinat examinării copiilor de la 3 ani la 7 ani şi 3 luni. Normele sunt aduse la zi, fiind adiţionate ansamblurile de obiecte.

În 1991, apare WISC III, o formă revizuită şi renormată a versiunii WISC-R.

Care sunt notele de distincţie a probei create de David Wechsler, care a făcut-o să deţină supremaţia în topul scalelor de inteligenţă? Am amintit deja că autorul nu a elaborat nici o probă originală (s-a inspirat în mare măsură din Binet-Simon şi din probele *Army Alfa*, create de Yerkes), ci a reevaluat deosebit de ingenios modul de combinare a testelor într-o scală.

- În primul rând, a împărţit în mod echilibrat probele între verbale şi nonverbale şi a lăsat probe de rezervă, ceea ce constituie o dovadă a pragmatismului american.
- În al doilea rând, a separat între o valoare globală şi două parţiale, în aprecierea rezultatelor (scalele de performanţă şi cele verbale au sisteme separate de etalonare).
- În al treilea rând, a îngustat la patru luni diviziunile cronologice de etalonare ale scalei.
- În al patrulea rând, stabilirea valorilor standard parcurge un drum mai lung. Dacă valoarea brută a probei se exprimă în secunde, acestea se transformă în punctaje care, la rândul lor, se folosesc la dobândirea valorilor standard.
- În al cincilea rând, David Wechsler a lansat un sistem foarte flexibil de stabilire a calităţii răspunsului. Se dau câteva exemple pentru acordarea de 0, 1 sau 2 puncte şi examinatorului i se oferă libertatea de a puncta şi alte răspunsuri considerate similare.

- În al şaselea rând, relaţia psihologului cu subiectul este activată prin posibilitatea intervenţiei cu ajutorul unor întrebări de precizare a răspunsului corect.
- În al şaptelea rând, protocoalele testului sunt foarte minuţios elaborate, în aşa fel încât cel mai dificil subiect supus examinării să fie deplin edificat în privinţa conţinutului solicitării. Aceasta dovedeşte o profundă înţelegere a situaţiei de examinare şi a blocajelor mintale ce pot interveni în cadrul ei. Formaţia de clinician a autorului îşi pune amprenta asupra concepţiei situaţiei de examinare.
- În al optulea rând, există un sistem de bonificaţie pentru performanţele de excepţie şi de atenuare a căderii bruşte a punctajului, în cazul unor răspunsuri imperfecte.
- În al nouălea rând, solicitările comprehensive sunt orientate spre situaţii cotidiene specifice societăţii şi instituţiilor americane. Discernământul este evaluat pentru situaţii foarte reprezentative.
- În al zecelea rând, la proba Wechsler este de apreciat umanismul concepţiei cu privire la examinarea psihologică. Testul prevede un întreg ansamblu de situaţii pentru ca subiectul să nu fie defavorizat şi să se prevină o situaţie de artefact. În privinţa calculării IQ-ului (care se exprima clasic prin raportul dintre vârsta mintală şi cea cronologică), Wechsler introduce următoarea formulă:

$$IQ = \frac{\textit{scorul realizat sau actual}}{\textit{media aşteptată a scorului pentru vârstă}}.$$

Această nouă formulă se bazează pe o supoziţie interesantă, stabilită sub forma unei axiome şi potrivit căreia valoarea IQ-ului rămâne constantă în raport cu vârsta normală, chiar dacă abilitatea intelectuală naturală se schimbă în plus sau în minus.

Variante foarte cunoscute ale scalelor Wechsler sunt scalele HAWIE (prescurtare de la *Hamburg-Wechsler-Inteligenztest für Erwachsene*) şi HAWIK (prescurtare de la *Hamburg-Wechsler-Inteligenztest für Kinder*). HAWIE este o prelucrare-standardizare germană, aparţinând lui A. Hardesty şi H.L. Lauber, apărută în anul 1955, pentru grupa de vârstă 10-59 de ani. Scala apare publicată împreună cu cartea lui Wechsler despre măsurarea inteligenţei, în care este definit conceptul de inteligenţă: „Inteligenţa este aptitudinea globală sau complexă a individului de a acţiona cu sens, de a se gândi la viitor, de a intra în relaţie cu mediul său" (*apud* Gregory, 1996, p. 229). În manualul scalei se face precizarea că noţiunea globală a inteligenţei nu poate fi ruptă de restul personalităţii. Un index valabil al aptitudinii globale de randament intelectual se poate obţine atunci când testul măsoară nu numai aptitudinile intelectuale specifice şi globale, ci şi factorii nonintelectuali ai inteligenţei. Wechsler susţine această teorie, bazându-se şi pe experienţa sa clinică, şi pe cercetările engleze de analiză factorială. În expunerea teoretică din cadrul HAWIE, se reliefează împotrivirea clară faţă de teoria factorilor (L.L. Thurstone) şi faţă de teoria factorilor multipli (Thorndike), menţionându-se fundamentul constituit din *teoria celor doi factori* elaborată de Ch. Spearman. Extinderea acestei teorii, prin noţiunea de *aptitudine globală*, la construcţia testului pleacă de la supoziţia potrivit căreia sarcinile individuale nu sunt încărcate doar cu factorul *g*, ci şi cu factori nonintelectuali care pot fi evaluaţi.

HAWIK este, practic, o copie a WISC la care s-au adăugat noi itemi. Transformarea în note brute se face luându-se ca puncte de reper media = 10 şi abaterea standard = 3. HAWIE-R apare în 1991, iar HAWIK-R în 1983.

O altă creaţie foarte cunoscută şi apreciată este *matricea progresivă Raven* (J.C. Raven, 1938), desemnată iniţial să măsoare factorul *g*, al lui Spearman. *Matricea progresivă Raven* (MPR) constituie un test esenţial pentru măsurarea inteligenţei generale (este un test nonverbal de raţionament inductiv la stimuli figurali) şi cunoaşte următoarele trei versiuni:

- Colored Progressive Matrices (5-11 ani);
- Standard Progressive Matrices (de la 6 ani până la adulţi);
- Advanced Progressive Matrices.

Deşi MPR este un index fidel şi valid al inteligenţei, acesta este un suport minor pentru test, ca măsură a factorului *g*. Analiza factorială de tip Spearman relevă doi sau trei factori, incluzând raţionamentul prin analogie şi completarea de structuri simple. MPR este folosită pentru testarea suplimentară a persoanelor cu probleme de auz, de limbaj sau cu handicap fizic. Revizuirile au fost făcute în 1983 şi 1986.

Culture Fair Intelligence Test (CFIT, Raymond B. Cattell, 1920) este un test pentru măsurarea inteligenţei nonverbale fluide, care încearcă să diminueze influenţa datorată formaţiei culturale. CFIT este proiectat pentru vârste de la 4 ani până la vârsta adultă şi apare în trei versiuni, fiecare cu câte două forme echivalente. Fiecare formă constă din 4 subteste: serii, clasificări, matrice şi condiţii. Fidelitatea CFIT este sporită şi scorurile corelează foarte bine cu alte probe de inteligenţă recunoscute. CFIT este un bun test de inteligenţă, însă se situează probabil la limita culturală a celor mai bune teste tradiţionale. Testul are nevoie de restandardizare şi de revizuire.

Cognitive Abilities Test (CogAT, Thorndike şi Hagen, 1978) este un instrument de investigaţie reprezentativ pentru categoriile de teste şcolare de nivel multiplu, de uz curent. Cele 8 subteste ale CogAT includ o baterie verbală, o baterie nonverbală şi o baterie cantitativă. Testul este co-normat cu două liste de achiziţii: Iowa Test All Basic Skills şi Test of Achievement and Proficiency.

Shipley Institute of Living Scale (SILS, Shipley, 1940) era propus, la origine, ca un index de deteriorare intelectuală. SILS este compus dintr-o secţiune de vocabular, constând din 40 de itemi cu alegere multiplă, şi o secţiune de 10 itemi de tipul completare de spaţii libere. Testul nu funcţionează bine ca un indice de organicitate, însă poate fi folosit la nevoie ca un procedeu de masă pentru examinarea inteligenţei generale.

În paralel cu scalele individuale, se dezvoltă în continuare testele de inteligenţă de grup care au făcut carieră în timpul primului război mondial şi a căror origine relativ modernă o găsim în tehnica de completare a unui text cu spaţii libere, pusă la punct de către H. Ebbinghaus, în 1890.

Testul colectiv diferă de cel individual în cinci privinţe (*cf.* Gregory, 1996, p. 222):

- alegere multiplă *versus* răspunsuri deschise;
- măsură obiectivă de scor *versus* măsură individualizată cu influenţe subiective;
- aplicare în masă *versus* planificare individuală;
- grup de standardizare mare *versus* eşantioane restrânse;
- administrare în grup *versus* administrare individuală.

Experienţa în domeniul examinării psihologice prin intermediul testului ne permite să mai adăugăm cel puţin trei direcţii clare de deosebire:

- manipulări instrumentale simple în sarcină *versus* manipulări complexe;
- sursă de informaţii diagnostice reduse *versus* surse complementare foarte mari;
- interacţiune redusă subiect – examinator *versus* interacţiune crescută.

4.2.5. *Testul nonverbal*

O lacună pronunţată în evoluţia domeniului nostru de interes a fost completată prin apariţia, încă din anul 1912, a probelor nonverbale sistematizate, după ce psihologul australian Porteus a fost însărcinat oficial cu depistarea deficienţelor mintale în instituţiile şcolare din Melbourne, prilej cu care el experimentează aceste probe, folosind în acelaşi timp şi o primă adaptare a versiunii Binet-Simon pentru Australia. În 1914, are ideea de a utiliza labirinturile şi, drept urmare, apare prima versiune a acestora pentru vârste între 5 şi 14 ani, obţinând o corelaţie de 0,70 cu rezultatele de la scala Binet-Simon.

În 1919, Porteus emigrează în SUA şi se stabileşte la Vineland, unde face mai multe modificări în proba sa, iar în 1937 realizează Vineland Revision, care cuprinde 12 labirinturi cu aplicare de la 3 ani până la vârsta adultă, făcând, pe baza acestora, calculul vârstei mintale. Creaţiile sale se regăsesc în scalele compozite ale măsurării inteligenţei, realizate de Grace Arthur şi David Wechsler.

În 1920, Kohs propune testul cuburilor, în scopul de a clarifica intervenţia factorului limbaj, folosind o probă constituită din 16 cuburi colorate cu dimensiunea de 2,5 cm, având patru feţe colorate în roşu, galben, albastru şi alb şi două colorări diagonale în roşu-alb şi galben-bleu. Proba cerea reproducerea unor figuri colorate, desenate pe un carton, prin folosirea unor astfel de cuburi, utilizând mai întâi un număr de 4 cuburi pentru figuri simple, 9 cuburi pentru probleme mai dificile şi apoi 16 cuburi pentru cele foarte dificile. În 1933, se realizează versiunea cea mai cunoscută. Grace Arthur reia probele lui Kohs, cu modificări în ordinea de prezentare. Cuburile se păstrează în actualitatea psihodiagnostică, fiind utilizate în numeroase teste compozite de inteligenţă.

O remarcabilă realizare în materie de teste nonverbale este şi creaţia lui Rudolf Pintner şi Donald Paterson, cunoscută sub numele de *scala Pintner-Paterson*, care s-a bucurat din 1931 şi la noi în ţară de o largă răspândire, prin intermediul cercetătorilor din Iaşi şi Cluj-Napoca (vezi *Testul Pintner-Paterson*, 1972). Scala de nivel iniţiată de autorii amintiţi anterior se recomandă mai ales pentru copiii provenind din medii socioculturale defavorizate, în privinţa posibilităţii de valorificare a potenţialului verbal. Testul oferă o gamă largă de solicitări nonverbale: relaţie spaţială, completarea unor spaţii libere, substituţii etc. Formele concrete de solicitare se intersectează cu alte creaţii de acest gen (Kohs, Bonnardell, Arnold Gesell, Charlotte Buhler, Irene Lézine). Folosirea sa este încă de actualitate.

4.2.6. *Testele de aptitudini*

În domeniul selecţiei şi orientării profesionale, sunt realizate un număr relativ mare de teste care au fost construite plecând de la o analiză a activităţii profesionale, definite în termeni de aptitudini şi de cunoştinţe primite. După primul război mondial, s-au construit teste creion-hârtie destinate a vedea în ce măsură subiectul stăpâneşte, fără a fi învăţat formal, noţiuni de fizică utile în anumite profesiuni, noţiuni cum ar fi principiul levierului sau regulile de transmitere a mişcării prin intermediul angrenajelor. Alte teste se apropie de testele de inteligenţă practică a scalelor de performanţă, dar conţinutul lor este mai mult tehnologic. În 1934, la Paris, D. Weinberg prezintă un test de montaj de obiecte, în care se solicită remontarea unor obiecte simple şi uzuale (un cleşte de rufe sau un zăvor) prezentate demontat. Alte teste simulează activitatea profesională. Cu testul Le Tourneu, J.M. Lahy imaginează o probă (în anul 1925) în care, prin acţionarea a două manivele, se deplasa un ştift pe un anumit traseu. Acest test a făcut carieră în domeniul selecţiei psihologice din domeniul transporturilor şi nu numai. Testele de acest gen tind să-şi piardă finalitatea iniţială – selecţia şi orientarea profesională – pentru a deveni tehnici de evaluare a psihomotricităţii.

Primul inventar de interese profesionale este prezentat în 1920, de către B. Moore, iar în 1927 apare prima versiune a inventarului lui Strong (Strong Vocational Interest Blank), inventar care, după multiple transformări şi perfecţionări, este foarte larg utilizat la ora actuală.

Testele de aptitudini măsoară mult mai specific şi delimitativ abilităţile umane decât testele de inteligenţă. În mod tradiţional, testele de inteligenţă evaluează un construct mult mai global decât inteligenţa generală. O aptitudine simplă vrea să măsoare o abilitate dintr-un domeniu, iar o baterie de aptitudini vrea să furnizeze scoruri în fiecare arie distinctivă de abilităţi.

Testele de aptitudini au apărut după cele de inteligenţă din două raţiuni: una statistică şi alta de natură socială. Nu în toate cazurile se delimitează între testele de aptitudini şi cele de inteligenţă, acestea din urmă fiind considerate o formă a celor de aptitudini. Problema statistică este aceea a unei noi tehnici, analiza factorială, care era utilizată pentru a discerne ce aptitudine este primară, deci distinctă de toate celelalte.

A doua raţiune a fost absenţa aplicaţiilor practice pentru o serie de instrumente rafinate. Abilităţile intelectuale puse în evidenţă de grupele testelor de inteligenţă folosite în primul război mondial nu erau suficiente pentru a selecta candidaţi adecvaţi pentru aviaţie, de exemplu. Forţele armate au rezolvat această problemă dezvoltând o baterie specializată de 20 de teste de aptitudini administrată la bărbaţii care treceau în prealabil de testele-filtru. Examinările cu astfel de teste erau folosite iniţial în selecţia piloţilor şi navigatorilor, pentru serviciile armatei.

Încă de la începuturi, au existat controverse în privinţa distincţiei între doi termeni: aptitudine şi capacitate. În opinia unor reputaţi autori din domeniul psihologiei industriale, „testele de aptitudini măsoară posibilitatea unui individ de a învăţa o muncă, dacă acesta beneficiază de formarea corespunzătoare, iar ele măsoară la fel calităţile sale latente"; „testele de capacitate sau de cunoştinţe achiziţionate măsoară nivelul individului în funcţia pentru care el este testat, precum şi cunoştinţele asupra acestei munci" (Tiffin şi McCormick, 1967, p. 34).

4.2.7. *Tehnicile proiective şi chestionarele de personalitate*

Metodologia psihometrică dezvoltată în legătură cu măsura eficienţei intelectuale va fi utilizată pentru a evalua şi alte aspecte ale conduitei. Aplicarea psihometriei la studiul personalităţii este bine conturată la începutul anilor '20.

Există în rândurile specialiştilor din domeniul psihodiagnosticului o serioasă rezervă cu privire la identificarea testului psihologic cu unele construcţii psihometrice destinate investigării personalităţii, deci cu privire la tratarea testului şi a construcţiilor ca realităţi aparţinând aceleiaşi familii a instrumentelor de investigaţie psihologică. Aceste reţineri sunt gradat reliefate, dar cele mai mari sunt în legătură cu aşa-numitele tehnici proiective şi cu chestionarele de personalitate. Totuşi, autorii unor astfel de tehnici îşi numesc creaţiile teste, la fel cum fac şi comentatorii lor (vezi, în acest sens, Rorschach, 2000, şi

Deri, 2000). Vom reveni asupra respectivelor distincţii, dar considerăm că este util să prezentăm şi creaţiile psihometrice din domeniul investigării personalităţii, cu atât mai mult cu cât avem de-a face cu eforturi constructive de un deosebit rafinament intelectual (ne referim cu prioritate la creaţiile lui Hermann Rorschach şi Lipot Szondi). Patentul testului proiectiv modern se consideră că-i aparţine psihiatrului elveţian Hermann Rorschach şi datează din anul 1920 (cu toate că, în anul 1904, Carl Gustav Jung pusese deja la punct şi folosise un test de asociaţie a cuvintelor), care creează o serie de 10 planşe cu stimuli ambiguu structuraţi şi elaborează o metodă adecvată de exploatare conceptuală a răspunsurilor la aceşti stimuli. Nevoia unui instrument adecvat cercetării clinice îşi pune amprenta asupra raţiunii apariţiei unor astfel de creaţii.

TAT (Thematic Apperception Test) al lui Morgan şi Murray (1935) a fost creat ca un instrument de studiu al personalităţii normale. Cei doi autori americani îşi manifestă creativitatea prin serii de fotografii care înfăţişează persoane angajate într-o interacţiune ambiguă.

Tot la serii de fotografii (de această dată individuale) apelează şi psihiatrul maghiar Lipot Szondi. Acesta aranjează 48 de fotografii în şase serii de câte opt, înfăţişând tipuri de pacienţi psihiatrici (homosexuali, epileptici, sadici, isterici, catatonici, paranoici, maniaci şi depresivi). Sistemul de examinare propus de Szondi se bucură de o anumită notorietate în Statele Unite, după folosirea testului de către Susan Deri.

Chestionarul de personalitate este privit ca un instrument prin excelenţă clinic, care oferă informaţii rezultate din introspecţia subiectului, prin folosirea unor întrebări privitoare la stările trăite de subiect în multiplele aspecte ale vieţii sale de relaţie. Valorile numerice rezultate după corectarea unui chestionar de personalitate au un caracter pronunţat convenţional, în comparaţie cu cele rezultate din testele de randament, de exemplu. Se vorbeşte, în cazul chestionarelor de personalitate, de o măsurare ipsativă, în care subiectul însuşi stabileşte dimensiunile valorice ale scalei. Primul inventar de personalitate (Personal Data Sheet) este prezentat de Robert Woodworth, în 1914, sub forma unei standardizări a unui examen psihiatric simplificat. Woodworth Personal Data Sheet apare în 1919 şi este construit din 116 întrebări la care subiecţii trebuie să răspundă prin „da" sau „nu" (considerat precursorul MMPI, Minnesota Multiphasic Personality Inventory). În 1955, este prezentată o formă dezvoltată a acestui chestionar (considerăm sinonime noţiunile de *chestionar de personalitate* şi *inventar de personalitate*). Nu discutăm acum lipsurile acestui chestionar (printre care evidentă este doza de credulitate psihometrică), socotind meritorie apariţia unei noi direcţii de diversificare a mijloacelor de investigare, motivată de nevoi din domeniul psihologiei clinice.

Această formă de chestionar este urmată de Thurstone Personality Schedule (Thurstone, 1931), de Bernreuther Personality Inventory (Bernreuther, 1931) şi de Allport-Vernon Study of Value (Allport şi Vernon, 1941).

Minnesota Multiphasic Personality Inventory (1940), aparţinând lui S.R. Hathaway şi J.C. McKinley, rămâne un instrument de investigaţie a personalităţii, de referinţă până în prezent. Strămoşi ai unor astfel de creaţii pot fi consideraţi Galton şi Kraepelin, prin metodele de asociere liberă elaborate la sfârşitul secolului al XIX-lea. Tot Galton este cunoscut ca precursor al tehnicilor proiective, urmărind observarea asocierilor intervenite în timp de patru secunde de la apariţia unui stimul cuvânt, precum şi categorizarea unei astfel de asociaţii.

4.2.8. *Inventarele de interese*

Preocuparea pentru diagnosticul şi consilierea persoanelor normale a dus la apariţia primelor inventare de interese. Începutul este făcut în 1912, de către Thorndike, care înregistrează tendinţele de dezvoltare a intereselor pe un lot de 100 de studenţi. În 1919-1920, Yookum creează 1.000 de itemi referitori la interesele din perioada dintre copilărie şi maturitatea incipientă. O parte dintre aceşti itemi au fost încorporaţi în Carnegie Interest Inventory. În 1926-1927, Cowdery îmbunătăţeşte şi rafinează inventarul lui Carnegie, sporind numărul de itemi şi comparând rezultatele celor trei grupe-criteriu (doctori, ingineri, avocaţi) cu grupe de control de nonprofesionişti, dezvoltând totodată o importantă formulă pentru itemi. Cowdery a fost primul psiholog care a realizat o validare încrucişată (*cross validation*). El testează noi scale pe grupuri adiţionale de doctori, ingineri şi avocaţi, pentru a argumenta că discriminările găsite în studiile originale erau mai degrabă fidele diferenţelor de grup decât capitalizării variaţiei erorii.

Edward K. Strong revizuieşte testul Cowdery şi alocă 36 de ani din viaţă pentru dezvoltarea soluţiilor empirice destinate modificării instrumentului cunoscut ca Strong Vocational Interest Blank (SVIB). Testul are soluţii de punctaj pentru fiecare sector ocupaţional. Versiunea modernă, Strong Interest Inventory, este folosită, în mare măsură, în consilierea profesională. În 1934, îşi face apariţia o realizare remarcabilă, competitor serios al SVIB-ului, şi anume Kuder Preference Record. Principala diferenţă faţă de inventarul lui Strong constă în alegeri forţate în rândul triadelor de itemi. Este un test ipsativ, comparând forţa relativă a intereselor în plan individual, şi nu răspunsurile la variate grupe profesionale. Revizia se numeşte Kuder Occupational Interest Survey (Kuder, 1966 ; Kuder şi Diamond, 1979, şi Zitowski, 1985, *apud* Gregory, 1996, p. 28).

4.2.9. *Testele educaţionale*

Ansamblul probelor din *Army Alpha* (care reprezintă intrarea în era testelor de grup) a fost preluat apoi în domeniul industrial şi în cel al şcolilor publice. Psihologii lui Yerkes şi-au continuat carierele în industrie şi educaţie (ei fondează noţiunea de *teste de inteligenţă de tip creion-hârtie*). Aplicaţiile practice cuprindeau examinări de admitere în facultate, de stabilire a achiziţiilor şcolare, de stabilire a nivelului aptitudinal. O consecinţă specifică a experienţei *Army Alfa* o constituie înfiinţarea National Research Council, o organizaţie guvernamentală a oamenilor de ştiinţă, care planifică National Intelligence Test, program în cadrul căruia urmau să fie examinaţi şapte milioane de copii, începând din anul 1920. Teste bine cunoscute din scala Wechsler, Scholastic Aptitude Test şi Graduate Record Examination, aveau rădăcini în cele oferite de Yerkes, Otis etc.

College Entrance Examination Board (CEEB) a fost constituit la întretăierea secolelor al XIX-lea şi XX pentru a ajuta la evitarea copierilor la examinarea candidaţilor din facultăţile americane. Examenele timpurii erau alcătuite sub forma unui scurt eseu, însă acestea au fost schimbate când C.C. Bringham a fost numit secretar al CEEB, imediat după primul război mondial. În 1925, College Board decide să construiască teste de aptitudini şcolare pentru admiterea la facultate.

Noile teste cuprind forme de examinare familiare, destinate completării propoziţiilor, analogiilor şi găsirii următorului număr într-o secvenţă. Sistemul de punctare a fost introdus în 1930, conferindu-le testelor obiective de grup o mult mai mare eficienţă

decât înainte. La început, denumirea testelor a fost Scholastic Aptitude Tests, iar apoi au fost cunoscute ca Scholastic Assessment Tests.

Funcţia CEEB a fost subsumată printr-o instituţie nonprofit, intitulată Educational Testing Service (ETS). ETS direcţionează dezvoltarea, standardizarea şi validarea câtorva teste bine cunoscute, ca Graduate Record Examination, The Low School Admission Test şi Peace Corp Entrance Tests.

Terman şi colaboratorii de la Stanford s-au preocupat de perfecţionarea standardizării testelor de achiziţie. Stanford Achievement Test (SAchT) a fost publicat pentru prima dată în 1923. O versiune modernă beneficiază de o largă utilizare şi în prezent. Încă de la început, SAchT cuprinde principii psihometrice moderne, cum ar fi normarea subtestelor, astfel încât variabilitatea intersubiecţi să poată fi evaluată şi selectată într-un foarte larg eşantion reprezentativ.

Psihometria se întinde dincolo de psihologie, mai ales în domeniul cunoştinţelor şcolare. Ca urmare a primelor lucrări ale lui Edward Lee Thorndike, metoda testelor este utilizată pentru evaluarea achiziţiilor şcolare. Pe baza acestei metodologii, se dezvoltă o critică a notării şcolare tradiţionale şi va apărea o nouă disciplină, *docimologia*, având drept obiect studiul sistematic al examenelor şi concursurilor, mai general al sistemelor de notare.

La sfârşitul anilor '30, exista deci o mare varietate de teste psihologice. În 1938, apare pentru prima dată un anuar comentat al testelor psihologice americane, *The Mental Measurement Yearbook*, care va fi periodic actualizat. În anul de apariţie sunt enumerate un număr de 331 de teste, dintre care o cincime sunt teste de inteligenţă (alte teste cuprind interesele, personalitatea şi, mai ales, cunoştinţele şcolare).

4.3. Evoluţia testelor după al doilea război mondial

După încheierea celui de-al doilea război mondial, până la sfârşitul anilor '70, domeniul testelor continuă să se dezvolte, chiar dacă nu se observă inovaţii majore, de tipul celor din perioada interbelică. Spre sfârşitul anilor '70 apar primele teste inspirate din teoria genetică a lui Piaget. Criticile făcute testelor devin mai frecvente, iar domeniul cunoaşte o slăbiciune care marchează mai ales utilizarea testelor colective. Încep să fie percepute posibilităţile oferite de tehnologiile informatice, iar psihologia cognitivă pare să fie capabilă să reînnoiască testele în conţinutul lor. Totuşi, persistenţa unui anumit deficit teoretic şi criticile asupra folosirii testelor conduc, începând din anii '70, la un relativ declin al folosirii testele.

Între 1939 şi 1945, psihologii participă la efortul de război al ţărilor lor prin punerea la punct a procedurilor de selecţie a militarilor. Participarea este mai mare decât în primul război mondial. Americanii îşi supun toţi soldaţii la aplicarea unei baterii de teste de inteligenţă, de aptitudini şi la un interviu. Între 1935 şi 1945, sunt examinaţi în acest fel patru milioane de recruţi. În anul 1941, britanicii adoptă o procedură similară, care va fi aplicată la sute de mii de persoane. Reich-ul a organizat un excelent serviciu psihologic de selecţie a militarilor, care a funcţionat până aproape de sfârşitul războiului... când nu prea mai avea ce selecta. Germania este prima ţară din lume care a elaborat o lege de funcţionare a serviciilor psihologice, în anul 1941. În Franţa, un

dispozitiv permiţând selecţia persoanelor din uzinele de armament este pus în funcţiune în 1939-1940. Henry Piéron coordonează serviciul de psihofiziologie şi de selecţie a personalului din aviaţie. Şi în alte ţări europene, psihologii participă la efortul de război, fără ca acesta să fie însă la fel de sistematizat.

La ieşirea din al doilea război mondial se observă o accentuare a difuzării testelor, fenomen asemănător cu cel de după primul război mondial. În SUA, practica testelor în sistemul şcolar se generalizează. Se mizează mult pe teste pentru a detecta talentele de care au nevoie economia şi administraţia (mai ales după lansarea primului satelit artificial al Pământului de către URSS, în 1957, fapt ce a relevat o întârziere a SUA în domeniul spaţial). Admiterea în universităţi şi în diverse instituţii de învăţământ superior tinde să se facă pe baza testelor psihologice, mai ales a SAT, care devine astfel un fel de etalon al nivelului populaţiei şcolare (scăderea scorurilor la SAT, observată între 1963 şi 1977, va fi atribuită, cel puţin din 1970, diminuării calităţii educaţiei). Folosirea bateriilor multifactoriale, inspirate din lucrările lui Thurstone, este din ce în ce mai frecventă în şcolile din învăţământul secundar, în industrie şi în comerţ. Aceste teste sunt mai bine adaptate la problemele de orientare şi selecţie decât testele unidimensionale de inteligenţă. Differential Aptitude Test, cu prima versiune apărută în 1947, este un bun exemplu de baterie multifactorială. Dar testele de inteligenţă generală sunt larg utilizate (în 1946, când C. Burt şi R.B. Cattell au ideea, bizară în aparenţă, de a constitui o asociaţie – MENSA – având drept prim obiectiv regruparea persoanelor al căror IQ depăşea nivelul existent în 98% din populaţie).

În Franţa, în perioada postbelică testele se dezvoltă cel mai rapid. În 1945, sunt create serviciile de psihologie şcolară, iar psihologilor şcolari li se atribuie sarcini de detectare a elevilor care prezintă handicap mintal şi a celor cu dificultăţi şcolare. Folosirea testelor psihologice în orientarea profesională a rămas relativ redusă în anii '20 şi '30, în ciuda eforturilor promotorilor lor. După cel de-al doilea război mondial, folosirea acestora va deveni aproape generală. Se pune în aplicare un tip de intervenţie caracterizat printr-un examen colectiv, prin intermediul testelor de grup, şi un examen individual, incluzând un interviu şi aplicarea unor teste. În industrie, dezvoltarea testelor este foarte rapidă, cu toate că anumite condiţii de validare a instrumentelor şi de formare a celor care le aplică şi le interpretează nu sunt întotdeauna îndeplinite.

Folosirea testelor de inteligenţă se dezvoltă în mod egal în America Latină şi în mai multe ţări din Asia (există deja realizări semnificative în ţări ca Argentina, Brazilia, Japonia şi Coreea de Sud). După război, se fac eforturi pentru a adapta testele la populaţia africană.

Alte forme de teste se dezvoltă mai rapid: în 1961, dintre cele 2.126 de teste disponibile la editorii americani, doar 11% erau teste de inteligenţă. Perioada de după război este caracterizată de creşterea numărului de aplicaţii ale psihologiei, în general (care existau şi înainte de război, dar într-un număr mai redus), dar mai ales de aplicaţii rezultate din psihologia socială, din mişcarea relaţiilor umane, apărută în anii '30 în SUA şi introdusă în Europa de după război, precum şi din domeniul psihologiei clinice. Practic, psihotehnica începează să deţină hegemonia în câmpul psihologiei aplicate. Dar noile aplicaţii ale psihologiei nu concurează cu testele: ele vizează rezolvarea altor probleme din domeniul ştiinţific, social şi economic.

4.3.1. *Dezvoltări conceptuale şi metodologice*

Dacă metodologia testelor mintale era deja bine pusă în aplicare în anii '30, ea progresează continuu în perioada imediată, de după al doilea război mondial până la începutul anilor '70. Apare teoria nivelurilor de măsură a lui S.S. Stevens (1946 şi 1951), care oferă un fundament pentru o serie de practici inadecvat asigurate, specificând regulile ce trebuie respectate şi postulatele ce trebuie să fie explicite în orice operaţie de măsură. Problemele relative la definirea dimensiunilor sunt, de asemenea, generalizate. L.J. Cronbach prezintă, în anul 1951, coeficientul alpha, care este actualmente considerat indicele cel mai general al omogenităţii unui ansamblu de itemi. Sunt elaborate mai multe modele în care aptitudinea vizată este considerată o variabilă latentă, funcţie a reuşitei la itemii propuşi. Metodele de analiză factorială progresează în mod continuu şi se diversifică. Sunt aduse soluţii care fac relativ indeterminate rezultatele analizelor: aceea a numărului de factori necesari să fie extraşi pentru a explica ansamblul corelaţiilor între variabile, aceea a alegerii unei soluţii dintre toate care sunt posibile.

Totuşi, aceste perfecţionări metodologice nu se traduc întotdeauna printr-o ameliorare a calităţii testelor prezentate publicului. Pentru SUA, Carroll (1982) notează că obiectivele lui *Mintal Measurement Yearbook*, care vizează explicit ameliorarea calităţii testelor, nu au fost decât parţial atinse şi că destul de multe probe nu satisfac normele elaborate de American Psychological Association şi American Educational Research Association (primele ediţii ale acestor norme apar în 1954). În majoritatea ţărilor, discrepanţele dintre dezvoltarea metodologiei şi realizările efective sunt încă mai mari.

Impactul lucrărilor lui Thurstone este foarte puternic, începând cu anii '40. Pentru mulţi psihologi problema principală nu constă în a preciza şi specifica măsurările inteligenţei generale, ci mai curând în a defini (prin tehnicile de analiză factorială) şi în a elabora forme diverse de inteligenţă. Cercetările întreprinse la ieşirea din cel de-al doilea război mondial şi în mare măsură „datorate" acestuia au permis evidenţierea a numeroase dimensiuni. Printre autorii care au propus o prezentare ordonată a acestor dimensiuni, trebuie să-i amintim pe J.P. Guilford şi R.B. Cattell. Aceşti autori au inventat teste, operaţionalizând diverse dimensiuni ale inteligenţei. J.P. Guilford, participant activ la selecţia piloţilor în timpul războiului, conduce timp de 21 de ani proiecte asupra „aptitudinilor de înalt nivel". În timpul derulării proiectelor, s-a realizat un inventar de aptitudini cognitive. În modelul său privind structura intelectului, prezentat pentru prima dată în 1956 şi expus în detaliu într-o lucrare din 1967, Guilford postulează 120 de dimensiuni ale inteligenţei, pe care le caracterizează în funcţie de trei parametri: tipul de material pe care se exersează activitatea mintală, tipul de operaţii solicitate şi tipul de finalitate a activităţii cognitive.

R.B. Cattell, psiholog englez şi activând în cea mai mare măsură în SUA, poate fi considerat reprezentant al şcolii engleze de psihologie diferenţială. Este, în primul rând, un psiholog al personalităţii şi se interesează de inteligenţă ca manifestare a personalităţii. Fidel tradiţiei engleze, el studiază modul de ierahizare a dimensiunilor inteligenţei. Spre sfârşitul anilor '60, prezintă un model cu şase mari dimensiuni (care nu sunt independente - un factor general se află în vârful ierarhiei), incluzând şi *inteligenţa fluidă*, capacitate de raţionament puţin legată de achiziţiile şcolare, şi *inteligenţa cristalizată*, strâns asociată cu aceste achiziţii. În aceeaşi perioadă, J.L. Horn prezintă un model foarte apropiat de acela a lui Cattell.

B.C.U. „M. EMINESCU" IAŞI

Deşi Cattell, Guilford şi Thurstone s-au străduit să demonstreze că există forme diverse de inteligenţă, rămân încă multe de făcut pentru a preciza natura acestora. Adesea, ele nu sunt definite decât prin comentarii asupra tipului de probleme care permit să fie reperate. S-au identificat puţine relaţii între studiul mecanismelor de funcţionare intelectuală şi acela al diverselor modalităţi de eficienţă a funcţionării respective.

Spre sfârşitul perioadei luate în vizor, când se înregistrează dezvoltări evidente ale psihologiei cognitive, legătura dintre psihologia generală şi cea diferenţială, ruptă în anii '20, se va restabili.

La sfârşitul anilor '60, există două psihologii cognitive bine individualizate. Prima ar fi psihologia europeană, în care Piaget este lider necontestat (Geneva fiind capitala unor astfel de preocupări) şi care îşi propune să studieze, pe parcursul unei serii de stadii, evoluţia structurilor gândirii. Această psihologie se dezvoltă încă de la începutul secolului XX şi se opune atomismului şi empirimului behaviorist. Pe de altă parte, apare psihologia americană, psihologia tratamentului infomaţiei, care, utilizând metaforele computerului, îşi fixează drept obiectiv analiza proceselor prin intermediul cărora informaţia este percepută, stocată şi utilizată pentru a elabora conduitele. Această psihologie a devenit, mai întâi în SUA, apoi în întreaga lume, paradigma dominantă[1]. Ea se opune behaviorismului prin focalizarea asupra proceselor interne şi alege un mod foarte analitic de abordare a fenomenelor psihologice. Aceste două psihologii, la ora actuală relativ bine coordonate, vor contribui la o reînnoire a bazelor teoretice ale testelor.

Ideea de a construi teste, plecând de la teoriile piagetiene, este relativ veche. La sfârşitul anilor '30, la Geneva ia amploare preocuparea de a standardiza examenul clinic piagetian, în scopul de a construi probe permiţând un diagnostic al nivelului de dezvoltare mintală mai bine fondat teoretic decât cel permis de probele de tip Binet-Simon. Barbel Inhelder (începând din 1943) a fost prima care a încercat să utilizeze probele piagetiene pentru diagnosticul raţionamentului la debilii mintal. Mai multe tentative de construcţie a unei scale piagetiene de dezvoltare cognitivă, vizând a concilia exigenţele contradictorii ale metodei „clinice" sau „critice" ale şcolii geneveze cu acelea ale standardizării psihometrice, au avut loc în anii '60, dar nici una nu a avut rezultate satisfăcătoare. O dată cu lucrările lui F. Longeot (1969), vor apărea şi primele teste extrase din teoria piagetiană: o serie de teste colective, „testele de operaţii formale", şi un test individual, *„échelle de développement de la pensée logique"* sau EPL. Obiectivul lui Longeot este practic şi teoretic. El doreşte să pună la dispoziţia psihologilor şcolari şi a consilierilor de orientare noi instrumente de observare. Reluând o ipoteză a lui M. Reuchlin (1964), îşi propune să stabilească o legătură între psihologia diferenţială şi psihologia dezvoltării a lui Jean Piaget, întrebându-se, mai ales, în ce măsură factorul *g* al diferenţialiştilor poate fi explicat prin noţiunea piagetiană de *structură de ansamblu, de stadiu*[2]. Alte teste de inspiraţie piagetiană au fost construite puţin mai târziu, în SUA, într-un mod cu totul independent şi într-o perspectivă destul de diferită. Se pune acum problema de a afla

1. O sinteză a acestei direcţii de cercetare a fost realizată de M. Miclea, „Psihologia cognitivă şi inteligenţa artificială", in I. Radu (coord.), *Introducere în psihologia contemporană*, Editura Sincron, Cluj-Napoca, 1991, pp. 199-215.
2. Cercetările care s-au derulat pe o perioadă de timp considerabilă (în această privinţă) sunt bine sistematizate în Serge Larivée, Sylvie Normandeau, Sophie Parent, „La filière francophone de la psychologie développementale différentielle", in *L'année psychologique*, 1996, pp. 291-342.

dacă testele de inteligenţă de inspiraţie psihometrică şi probele piagetiene evaluează aceeaşi inteligenţă sau două forme de inteligenţă diferite. Scala individuală a lui Longeot, ce furnizează, desigur, informaţii mai bogate decât testele colective, nu a avut succesul pe care îl merita. Testele rezultate din teoria piagetiană au contribuit, de asemenea, la evoluţia acestei teorii, punând în evidenţă puternica eterogenitate intraindividuală a nivelului de dezvoltare, eterogenitate care face contestabilă ideea şi, prin urmare, şi măsurarea unui nivel general de dezvoltare. Cercetătorii, autori ai unor astfel de aplicaţii, au fost la originea unui curent de cercetare original, denumit adesea *psihologia diferenţială a dezvoltării.*

Întâlnirea dintre teste şi psihologia tratamentului informaţiei a avut loc mai târziu şi a trebuit să se aştepte anii '70 ca să apară unele lucrări semnificative. De asemenea, începând din 1970 efectele informaticii asupra testelor vor fi sensibil mai pronunţate.

4.3.2. *Dezvoltări ale testelor de la începutul anilor '70 până în prezent*

Testele sunt, după acest reper temporal, din ce în ce mai frecvent analizate şi construite plecând de la conceptele psihologiei cognitive a tratamentului informaţiei. Posibilităţile oferite de informatică sunt din ce în ce mai evidente.

Creşterea numărului de investigaţii experimentale şi de elaborări teoretice din psihologia cognitivă influenţează în mod esenţial şi testele de inteligenţă. Deşi se susţine că psihologia cognitivă a lui Jean Piaget a fost la originea noilor teste, merită amintit aportul psihologiei tratamentului informaţiei, care este cel mai recent. Primele realizări semnificative datează din anii '70, o dată cu apariţia lucrărilor lui S. Sternberg asupra raţionamentului analogic (1977) şi ale lui E. Hunt asupra tratamentului verbal (1978). De atunci numeroşi cercetători au elaborat idei cu privire la contribuţia rezultatelor psihologiei cognitive la mai buna înţelegere a ceea ce măsoară testele clasice, precum şi cu privire la aportul în construcţia testelor cu un fundament teoretic mai consistent. Se pune în evidenţă rolul jucat de reprezentarea problemei, prin operaţiile mintale, conducând la modificările acestor reprezentări, şi prin procesele metacognitive, permiţând planificarea şi controlul activităţii de rezolvare. Majoritatea testelor clasice evaluează toţi factorii care contribuie la eficienţa unui test. De fapt, aceasta este raţiunea pentru care ele permit prognosticări satisfăcătoare. Evaluarea fiind însă globală, nediferenţiată, ea nu furnizează decât informaţii sărace despre dificultăţile cu care se confruntă subiecţii şi despre cauzele eşecurilor lor, şi deci despre mijloacele de remediere a construcţiei lor.

Construirea noilor teste pe baza realizărilor teoretice ale psihologiei cognitive a trezit mari speranţe. Printr-un demers similar celui al psihologilor de la sfârşitul secolului al XIX-lea, s-au trasat paradigmele experimentale din psihologia cognitivă, în scopul de a măsura eficienţa sau, mai adesea, viteza de execuţie, de operare mintală. Testele elaborate în această perspectivă nu sunt la fel de interesante cum s-a prevăzut. Desigur, măsurile obişnuite erau satisfăcătoare din punctul de vedere al fidelităţii şi al validităţii lor teoretice (ele fac parte dintre modelele probate de funcţionare cognitivă). Dar, dat fiind caracterul lor prea elementar, se resimţea lipsa pertinenţei pentru analiza fenomenelor de învăţare în contextul social obişnuit. Eşecul acestor tentative poate fi relativizat. Pentru anumite tipuri de probleme bine circumscrise, nu mai este atunci vorba despre inteligenţă (la propriu), iar abordarea analitică este mai interesantă. Pentru învăţarea

lecturii, de exemplu, există baterii de probe permiţând să se identifice în mod precis unele cauze ale dificultăţilor întâlnite de către cei care învaţă.

În scopul de a depăşi limitele unei abordări prea analitice a fenomenelor psihologice, s-au construit câteva teste inspirate din modelele elaborate de neuropsihologia cognitivă. Această disciplină este relativ puţin direcţionată spre analiza fină a tratamentului informaţiei, ea căutând mai curând să definească marile funcţiuni şi să studieze interacţiunile lor. Dacă testele rezultate din acest curent de cercetare, de exemplu testul lui Kaufman, sunt mai bine fondate teoretic decât testele clasice, se poate spune că ele nu sunt radical diferite şi împrumută destul de numeroase situaţii. La începutul anilor '70, se aştepta ca psihologia cognitivă să reînnoiască în profunzime testele. Dar, până acum, procesul de reînnoire a fost limitat.

4.3.3. *Evoluţia practicii testelor*

Se cuvine să diferenţiem două feluri de practici: practica extensivă, care cuprinde aplicaţii ale testelor de grup în plan social, în domeniul învăţământului şi al muncii, şi practica testelor individuale în cadrul examenului psihologic. Este, de asemenea, necesar să ţinem cont de contextele naţionale.

Testele de grup folosite în prezent în Statele Unite se aseamănă mult. Toate cuprind itemi de aceeaşi natură (vocabular, informaţii generale – aceşti itemi sunt cei mai sensibili la *biais*, din ce în ce mai puţin frecvenţi –, raţionament cu privire la datele verbale numerice sau figurale) şi permit să se obţină un scor global general, exprimat în IQ standard. Practica acestor teste este în scădere, în sistemul educativ, fapt datorat criticilor repetate cărora le-au făcut obiectul şi restricţiilor în utilizarea lor, impuse de autorităţile judiciare şi academice, pe care le vom semnala într-un capitol ulterior. Practica testelor de grup rămâne totuşi foarte frecventă. Există mai multe explicaţii pentru această „rezistenţă la teste". Într-un sistem educaţional saturat de examene naţionale, testele îndeplinesc funcţia de reglare şi a le suprima presupune înlocuirea cu o altă procedură. În fapt, se propune, cel mai adesea, înlocuirea testelor de inteligenţă cu alte teste. De altfel, partizanii testelor au semnalat că aceste instrumente permit bune prognosticuri, atât cu privire la adaptarea şcolară, cât şi la cea profesională.

În Franţa, situaţia stă cu totul altfel. Cum am arătat mai sus, selecţia şcolară răspunde la notele date de profesori. Testele de grup (testele compozite şi testele factoriale) au fost totuşi foarte utilizate în învăţământul secundar, în anii 1960-1975, de către consilierii de orientare şcolară şi profesională. Se pune acum problema de a oferi posibilitatea profesorilor să facă observaţii diferite de acelea pe care le primesc în mod obişnuit. În cadrul politicii voluntariste de prelungire a şcolarităţii, se propune, de asemenea, să se repereze elevii a căror reuşită şcolară era slabă sau medie, dar ale căror rezultate la teste dau posibilitatea de a prognostica o reuşită şcolară satisfăcătoare în învăţământul superior. Colectivele însărcinate cu orientarea elevilor dispuneau de aceste informaţii şi erau libere să le utilizeze sau nu. Utilizarea masivă a testelor a provocat neînţelegeri fundamentale. Din punctul de vedere al psihologilor, informaţia pertinentă era discordantă cu evaluarea şcolară, dar mulţi profesori nu o reţin decât pe aceea care era în acord cu asemenea evaluări. Acest mod de utilizare a testelor conduce atunci la legitimarea clasamentelor şcolare şi explică de ce testele erau, în general, acceptate de către profesori. Practica testării colective s-a diminuat considerabil şi, uneori, chiar a dispărut cu totul. Această

evoluţie amplă şi rapidă se explică în mai multe moduri : reacţiile critice faţă de teste au tulburat un anumit număr de psihologi. S-a luat cunoştinţă de faptul că testele colective furnizează o informaţie adesea prea săracă pentru a fi cu adevărat utile profesorilor, şi anume atunci când orientarea devine puţin selectivă, iar cercetarea înzestrărilor îşi pierde oarecum din interes. De asemenea, nu trebuie neglijat rolul psihologului şcolar în învăţământul secundar, care a evoluat considerabil : iniţial, rolul considerat clasic era acela de a stabili diagnosticul pentru a realiza consilierea ; actualmente îi revine în mod esenţial rolul de a ajuta la dezvoltarea personală.

În domeniul recrutării profesionale, testele de inteligenţă sau de aptitudini sunt utilizate frecvent în Statele Unite şi, inegal, după ţări, în Europa. Destul de adesea aplicate în Franţa, Anglia, Belgia, ele sunt mult mai puţin folosite în Germania (C.L. Leboyer, 1994).

Este dificil de judecat evoluţia practicii, în ceea ce priveşte testele individuale. În Franţa, frecvenţa utilizării lor a fost relativ puţin afectată de diversele critici, care au vizat testul în general. Testele sunt rar utilizate în orientarea şcolară şi profesională, dar folosirea lor este totdeauna frecventă în domeniul psihologiei şcolare şi al sănătăţii. Apar nevoi noi în materie de diagnostic al caracteristicilor individuale cognitive. În domeniul inserţiei tinerilor în situaţii de eşec şcolar, de exemplu, numeroşi formatori doresc să dispună de date psihologice care să le indice pe ce competenţe pot să se sprijine pentru a defini obiectivele şi metodele de formare. Una dintre frânele în folosirea testelor individuale rezidă în condiţiile de muncă ale psihologilor, care, adesea, sunt constrânşi de a lua în sarcină un număr prea mare de subiecţi.

Există reviste ştiinţifice care se ocupă de elaborarea unor viziuni prospective asupra testelor şi asupra utilizării lor (de exemplu, în numărul special al revistei *Pratiques psychologiques*, publicat în 1996, având ca temă testele şi utilizarea lor).

M.L. Bruschon-Schweitzer, în 1991, afirmă că în Franţa testele de inteligenţă sunt utilizate sistematic de către 31 % şi, ocazional, de către 32 % dintre serviciile de recrutare ale întreprinderilor şi ale cabinetelor de selecţie.

4.3.4. *Intensiv şi extensiv în evoluţia testului*

Un bun cunoscător al testelor psihologice, care parcurge forme de teste psihologice apărute după 1940, are senzaţia de *déjà vu*. Formele fundamentale sub care se prezintă solicitarea în testul psihologic au fost create, în mare parte, până în 1940, ulterior având loc modificări, adaptări, perfecţionări, combinări ale subtestelor construite cu un anumit scop. În anii '50, o dată cu apariţia cărţilor lui Lord şi Novick, pe de o parte, şi Gulliksen, pe de altă parte, testul psihologic clasic este bine fundamentat în teorie. În acest caz, se poate vorbi despre o stagnare a dezvoltării testului? În nici un caz, considerăm noi, dar evoluţia se manifestă cu prioritate în domeniul intensiv.

Este necesar să fie făcută o precizare, în privinţa evoluţiei testului psihologic. Am acordat deja un spaţiu considerabil clasicilor probelor psihologice şi se pune problema dacă nu cumva testul psihologic a rămas „încremenit" în formele de la începutul secolului. O privire superficială ar favoriza această primă impresie, mai ales că Stanford-Binet este o prezenţă de primă valoare în psihodiagnosticul contemporan, iar testele Wechsler sunt considerate cele mai reputate probe în examinarea inteligenţei, deşi sunt mai bine de 60 de ani de la apariţia primelor variante. Continuitatea folosirii unor probe în examinarea psihologică nu este o dovadă de stagnare, ci de certitudine valorică. Trebuie să avem în

vedere faptul că aceste probe se menţin prin efectuarea unui număr de revizuiri consistente şi de durată. Existenţa unor elemente metodologice perene, în examinarea psihologică prin intermediul testelor psihologice, este o dovadă de continuitate într-o ştiinţă atât de diversificată conceptual cum este psihologia. Variantele moderne ale probelor psihologice clasice nu mai sunt ceea ce au fost la primele ediţii, ci veritabile colecţii de componente verificate de practica profesională. În metodologia psihologică, sintetizatorii, integratorii informaţionali au fost cei care s-au menţinut în actualitate, pe criteriul meritului.

Testul psihologic cunoaşte două căi de evoluţie, aflate în strânsă interdependenţă : una extensivă şi cealaltă intensivă. Calea extensivă se manifestă prin următoarele caracteristici :

- lărgirea ariei geografice de răspândire ;
- deplasarea vârstei de aplicare la cele două limite, inferioară şi superioară ;
- diversificarea formelor sub care se prezintă solicitarea ;
- creşterea numărului domeniilor de aplicaţie ;
- mărirea numărului de situaţii în care se aplică, fapt ce necesită adaptări ;
- extinderea categoriilor de populaţie care beneficiază de acest fel de examinare psihologică.

Modalitatea intensivă are următoarele direcţii :

- creşterea obiectivităţii în evaluare (o instrucţie mai adecvată în evaluarea sarcinilor, înlăturarea dezechilibrelor în sarcinile din mai multe părţi, când avem de-a face cu două până la trei rezolvări parţiale, una şi nici o rezolvare parţială) ;
- perfecţionarea teoriilor care stau la baza construcţiei testelor şi apariţia unor concepţii noi ;
- puternica intervenţie, în construcţie şi exploatare, a statisticii (în special a celei inferenţiale) ;
- perfecţionarea analizei cantitative şi calitative a itemilor ;
- creşterea exigenţei cu privire la calitatea măsurării.

Testul psihologic a fost „împins" la diversificare de inadecvarea evidentă pentru o serie de persoane şi situaţii de examinare şi de exploziile sporadice ale cerinţelor de examinare psihologică. Lipsa de pregătire a celor ce le-au construit şi exploatat din punct de vedere diagnostic pentru a face faţă cerinţelor temporare foarte mari a dus, inevitabil, la suspiciuni şi serioase rezerve faţă de testul psihologic.

Efortul nostru a fost acela de a evidenţia succint o serie de „prototipuri" ale testelor psihologice, şi nu de a face o istorie amănunţită a acestora. Un asemenea demers ştiinţific ar fi gigantic, având în vedere numărul foarte mare de teste existente. Doar în domeniul cercetării dotărilor speciale se folosesc câteva mii de teste. Bibliografia destinată testelor psihologice, alcătuită de G.H. Hildreth în anul 1939, cuprindea deja 4.279 de teste şi scale de evaluare. În anul 1946, numărul acestora, raportat de acelaşi autor, depăşeşte 5.000 (*apud* Hofstätter, 1971, p. 289). Roger Mucchielli afirmă, într-o lucrare destinată examenului psihotehnic, din 1991, că la acea dată în lume existau peste 14.000 de teste (Mucchielli, 1992, p. 37).

Prezentări sistematice ale testelor psihologice aflate în prezent în exploatare au fost realizate de către firmele comerciale de profil, care funcţionează, de mai bine de 30 de ani, în peste 20 de mari centre urbane. Astfel, ECPA (Les Éditions du Centre de Psychologie

Appliquée), funcţionând de 50 de ani, a editat şi difuzat mai mult de 200 de teste. ECPA posedă un serviciu propriu de cercetare şi dezvoltare, dezvoltă activităţi permanente pentru psihologii practicieni, oferă asistenţă utilizatorilor produselor achiziţionate şi editează o revistă de psihologie aplicată (*La revue européenne de psychologie appliquée*). Testele produse sunt grupate în:

- teste de inteligenţă, cuprinzând probe individuale şi colective, utilizate pentru examinarea adulţilor, adolescenţilor, copiilor, precum şi un *baby-test*;
- teste de cunoştinţe;
- probe clinice;
- teste de aptitudini, înglobând teste de aptitudini intelectuale, manuale şi pentru profesiuni determinate;
- chestionare şi inventare de personalitate;
- teste obiective de personalitate;
- tehnici proiective.

Prezentarea fiecărui test cuprinde descrierea materialelor folosite, modul de administrare şi referiri la actualitatea criteriilor de valabilitate ale probei[1].

În alte cazuri, inventarele celor mai importante teste aflate în exploatare psihodiagnostică sunt prezentate de personalităţi ştiinţifice din domeniu şi oferă informaţii condensate asupra ultimelor noutăţi de specialitate (*cf.* Brickenkamp, 1997). Sistemelor computerizate de examinare prin intermediul testelor psihologice li se oferă o secţiune specială a lucrării noastre.

1. Pentru documentare am consultat catalogul ECPA din ianuarie 1997.

Capitolul II

TEORIA TESTULUI PSIHOLOGIC

În orice demers ştiinţific pertinent, există obligaţia precizării stricte a termenilor uzitaţi, operaţie care implică detaşarea denotaţiilor de conotaţii. Definirea adecvată a termenilor se impune mai ales în cadrul planurilor de cercetare de conţinut.

Noţiunea de test nu este folosită doar în psihologie, ci şi într-o serie de alte discipline. Putem spune fără riscul de a greşi că testul este un termen nespecific psihologiei. Se vorbeşte despre teste fizice, biologice, chimice etc. De asemenea, se vorbeşte despre teste medicale. Nu în rare cazuri, auzim de testarea motoarelor, a instalaţiilor şi, în teoriile ştiinţifice, de testarea unor modele. Consultarea unor dicţionare explicative generale, nespecializate în psihologie, indică diferenţieri ale domeniilor de utilizare a termenului „test"[1].

În sensul cel mai general, cuvântul „test" înseamnă procedeu de verificare care apelează frecvent la forma chestionarului. Este mai ales cazul probelor de verificare a cunoştinţelor din diverse discipline ştiinţifice sau preocupări umane sistematizate.

Menţiuni suplimentare sunt necesare pentru sensurile testelor medicale, biologice şi statistice. În cazul testului medical şi biologic, sensul este acela de probă care poate să pună în aplicare tehnici medicale variate (reacţii chimice sau biologice, prelucrări) şi să furnizeze o indicaţie pentru diagnostic (de asemenea, materialele folosite pentru aceste operaţiuni poartă numele de teste).

Testul statistic este o procedură matematică pentru verificarea ipotezelor ştiinţifice, constând din calculul unei valori numerice şi decizia (pe baza unei reguli fixate în prealabil) de acceptare sau de respingere a unei ipoteze nule (cu o anumită probabilitate de a lua o decizie incorectă, când ipoteza nulă H_0 este confruntată cu ipoteza alternativă H_1 – vezi Trebici, coord., 1985, p. 382).

Aplicarea unui test statistic presupune parcurgerea mai multor etape riguros realizate. Dintre zecile de teste statistice existente, amintim testul de concordanţă al lui Kendall, testul neparametric Mann-Whitney-Wilcoxon, testul T, testul Z etc.

1. Am consultat, în această privinţă: *Dicţionarul explicativ al limbii române*, Editura Academiei, 1984; *Dictionnaire de la langue française*, Hachette, Paris, 1994; *Le Petit Larousse*, Larousse, Paris, 1992; *Webster's Encyclopedic Unabridged Dictionary of English Language*, Grammercy Books, New York, 1996; *Cambridge International Dictionary of English*, Cambridge University Press, Cambridge, 1996.

1. Definiţii ale testului psihologic

Definirea testului psihologic de către specialiştii teoriei testelor diferă mai puţin în privinţa terminologiei utilizate, cât mai ales prin amploarea sau îngustimea referinţelor.

O definiţie de referinţă este cea a lui Gustav Lienert, autorul unuia dintre cele mai cunoscute tratate de teorie a testului psihologic clasic: „Un test este un procedeu ştiinţific de rutină pentru cercetarea uneia sau mai multor caracteristici delimitabile empiric ale personalităţii, în scopul emiterii unei afirmaţii (aserţiuni), pe cât posibil cantitative, despre gradul relativ al reliefării caracteristicii individuale" (Lienert, 1967, p. 7).

Într-un dicţionar de psihologie, Wilhelm Hehlmann defineşte testul ca „un examen simplu de stări, de fapte psihice sau fizice" (Hehlmann, 1968, p. 560). O definiţie mai generală nu credem că este posibilă, chiar dacă autorul exemplifică ulterior cu denumiri de aptitudini, talente, cunoştinţe pentru detectarea cărora se apelează la test.

Asociaţia Internaţională de Psihotehnică dă următoarea definiţie testului: „Este o probă definită, implicând o sarcină de îndeplinit, identică pentru toţi subiecţii examinaţi, cu o tehnică precisă pentru aprecierea succesului şi a eşecului sau pentru notaţia numerică a reuşitei. Sarcina poate să presupună o punere în aplicare fie a cunoştinţelor achiziţionate (teste pedagogice), fie a funcţiilor senzorial-motrice sau mintale (teste psihologice)" (Lussato, 1998, p. 7). Se observă, din această definiţie, excluderea din rândul testelor a construcţiilor psihometrice destinate investigării personalităţii. Astfel de puncte de vedere nu sunt singulare.

Arlette şi Roger Mucchielli consideră testul psihologic „o probă semnificativă (validată şi etalonată) pentru a evalua (prin raportare la rezultatele obţinute de alţii) de ce este capabil un subiect într-o situaţie dată sau caracteristicile personalităţii sale" (Mucchielli, 1992). L.J. Cronbach definea testul psihologic ca fiind „un procedeu sistematic pentru observarea comportamentului unei persoane şi a cărui descriere se face pe baza unei scale numerice sau a unui sistem categorial" (Cronbach, 1970, p. 26). O definiţie lapidară, cu referinţă schimbată faţă de cele enumerate anterior, îi aparţine lui Helmuth Lukesh: „Un test poate să fie, în sensul cel mai general, un eşantion de comportamente pe baza cărora se prognozează alte comportamente" (Lukesh, 1968, p. 560). În aceeaşi notă particulară se menţine definiţia dată de Lothar Michel şi Wolfgang Conrad: „Testul psihologic poate să fie desemnat ca un experiment psihologic specific, care serveşte recunoaşterii şi descrierii caracteristicilor psihice individuale" (Michel şi Conrad, 1982, p. 1). Definiţia dată de Lothar Michel şi Wolfgang Conrad pare să-i nege testului psihologic calitatea de metodă autonomă în cadrul psihologiei. Ioan Radu împărtăşeşte această opinie atunci când afirmă: „În primă aproximaţie, testul psihologic este un gen de experiment standardizat" (Radu, 1991, p. 44). Aceeaşi optică pare să o adopte şi Monica Albu, la începutul unei lucrări de specialitate consacrate construirii şi utilizării testelor psihologice: „Testul este un instrument al metodei experimentale, folosit cu precădere în investigaţiile cu caracter aplicativ ale psihologiei (testul psihologic), pedagogiei (testul pedagogic, testul şcolar), sociologiei (testul sociometric) şi biologiei (testul fiziologic)" (Albu, 1998, p. 13). Din această aserţiune, rezultă clar caracterul nespecific pentru psihologie al noţiunii de test (autoarea face precizarea că nu a încercat să formuleze o definiţie).

O definiţie succintă a testului psihologic aparţine Ursulei Şchiopu: „Probă, instrument de diagnosticare de caracteristici psihice" (1997, p. 690). Maurice Reuchlin oferă

o definiţie care surprinde aspecte esenţiale ale testului psihologic: „Probă, utilizată mai ales în psihologia diferenţială, care permite descrierea comportamentului unui subiect într-o situaţie definită cu precizie («consemnul» testului) prin referinţă la comportamentul unui grup definit de subiecţi plasaţi în aceeaşi situaţie"[1]. O serie de teoreticieni din domeniul teoriei testelor nu se mai străduiesc să elaboreze în lucrările lor o definiţie a testului, considerând, probabil, că lectorul este iniţiat în preocupările vizând perfecţionarea testului psihologic şi precizările noţionale specifice acestuia. O definiţie de largă circulaţie a testului psihologic este dată de Pierre Pichot, folosind denumirea originală cu care a fost lansat testul: „Se numeşte test mintal o situaţie experimentală standardizată servind drept stimul pentru un comportament. Acest comportament este evaluat printr-o comparaţie statistică cu acela al altor indivizi plasaţi în aceeaşi situaţie, permiţând astfel clasarea subiectului examinat, fie cantitativ, fie tipologic" (Pichot, 1994, p. 5).

Unii analişti ai teoriei şi practicii testului psihologic consideră că existenţa mai multor definiţii ale testului (aproape pentru fiecare autor câte una) este o dovadă de ambiguitate în acest domeniu, de deficienţe în conturarea clară a noţiunii respective. Or, rostul definiţiei este tocmai acela de a clarifica, de a înlătura confuziile, de a desprinde trăsături distincte, proprii numai respectivei realităţi. Logicianul Gheorghe Enescu desemnează definiţia ca fiind „o propoziţie sau un ansamblu de propoziţii prin care se determină semnificaţia unui termen ori se indică note caracteristice prin care o noţiune se deosebeşte de altele ori se indică modul de construcţie a unei clase de obiecte formale (ori secvenţe finite de asemenea obiecte)" (Enescu, 1985, p. 70).

Analizând domeniul teoriei definiţiilor, nu considerăm că această pluralitate este un fapt cu semnificaţie negativă, ce ar afecta calitatea abordării ştiinţifice a unui domeniu. Nimeni nu este vinovat pentru că există realităţi cu mai multe faţete şi, drept urmare, sunt posibile mai multe modalităţi pertinente de a le defini. Există metode diferite de a defini noţiunile ştiinţifice, a căror eficacitate pentru metodica cercetării empirice trebuie să fie cunoscută. Este importantă cunoaşterea anumitor reguli şi greşeli specifice care trebuie respectate, respectiv evitate, în definirea ştiinţifică. O definiţie adaptată nivelului de cunoştinţe, respectiv procesului de comunicare, poate să restrângă neînţelegerile şi să ducă la o alegere adaptată a metodelor pentru verificarea ipotezelor în procesul de decizie ştiinţifică. Demersul nostru explicativ este unul minimal, deoarece teoria definiţiei constituie o preocupare milenară a disciplinelor filosofice şi matematice.

O definiţie are următoarea structură de bază: A = df BC, în care *A* este entitatea care se defineşte, având consacrată denumirea latină *definiendum*, *df* desemnează relaţia de definiţie, iar *BC* - definitorul, având denumirea latină consacrată *definiens*. La rândul său, definitorul, cel care defineşte, este format din două părţi: B, constituind genul proxim, genul cel mai apropiat din care face parte A, şi C, constituind diferenţa specifică, adică acea proprietate care delimitează pe A în B. Logicianul român amintit anterior consideră că genul proxim este foarte dificil sau chiar imposibil de găsit, motiv pentru care se apreciază că este oportună indicarea unui gen apropiat. Funcţiunile unei definiţii sunt enumerate în continuare:

- identificarea obiectelor (precizare pozitivă);
- delimitarea obiectelor neidentificate (precizare negativă);
- precizarea semnificaţiei;

1. Vezi *Grand dictionnaire de la psychologie*, Larousse, Paris, 1992, p. 782. Definiţia din dicţionar îi aparţine lui Maurice Reuchlin, membru al colectivului de autori care l-a elaborat.

- realizarea de afirmaţii despre relaţiile dintre semne şi desemnare;
- sistematizarea cunoştinţelor;
- realizarea comunicării.

O definiţie poate să se realizeze într-un limbaj natural, într-un limbaj de specialitate, posibil într-un limbaj formal sau prin combinarea acestora. Formele de definiţie alese şi limbajul de definire ales trebuie să fie adaptate nivelului de cunoştinţe din domeniul obiectului definit. Aceasta înseamnă că nu este recomandată o *supradefinire*, obiectului fiindu-i adăugate caracteristici care nu sunt încă dovedite, sau o *subdefinire*, presupunând o renunţare la informaţii. Supradefinirea este apreciată critic, deoarece, în cazul existenţei acesteia, ipotezele (încă nedemonstrate) pot fi înţelese ca teze şi fapte (deja demonstrate).

De asemenea, forma aleasă a definiţiei trebuie să fie adaptată valorii funcţionale a unei definiţii. De exemplu, într-un proces de învăţare (instruire), ea trebuie să fie adaptată nivelului de cunoştinţe şi nivelului de înţelegere al celor care învaţă. Înseamnă că unul şi acelaşi lucru poate fi exprimat adecvat prin diverse forme de definiţie, fiecare formă aleasă prezentând compromisul a două condiţii: *atât de exact cât este posibil* şi *atât de exact cât este necesar*. Ca atare, o definiţie dintr-un tratat de specialitate poate diferi de cea aparţinând unui curs introductiv, cu referire la acelaşi obiect de interes. În această diferenţiere de exigenţe trebuie căutată cauza existenţei unei mari diversităţi de definiţii pentru una şi aceeaşi stare de fapt. Altă explicaţie constă în aceea că diverse forme de definiţie fac posibilă reliefarea diverselor aspecte ale *definiendum*-ului în *definiens*. În concluzie, pot fi acceptate simultan mai multe definiţii.

Parcurgând definiţiile cunoscute şi analizând deficienţele acestora, am elaborat următoarea definiţie pentru testul psihologic: „Procedeu standardizat de investigaţie psihodiagnostică, punând în evidenţă, prin diferite modalităţi de stimulare a subiecţilor, comportamente relevant diagnostice, exploatabile informaţional, pe baza comparării rezultatelor obţinute cu cele aparţinând unor eşantioane reprezentative de persoane aflate într-o situaţie de examinare identică şi care foloseşte, în descrierea şi explicarea concluziilor, termeni ştiinţifici specifici teoriei pe care se bazează construcţia sa".

Pe parcursul lucrării noastre, vom reveni asupra unor forme de definire utilizate în conceperea unui test psihologic.

2. Părţile componente ale unui test psihologic

În unele manuale, atunci când se prezintă succint un test, este expusă, prin fotografiere sau schiţare, doar partea de stimulare, însoţită sau nu de eventuale comentarii. În calitate de procedeu standardizat de investigaţie, un test psihologic are o serie de componente care, în ansamblu, asigură funcţionalitatea sa:

a) *componenta statică*, materială, a testului, denumită uneori şi partea de rechizite a testului, se subdivide în:
   - manualul testului, în care sunt prezentate descrierea, geneza constructivă, fundamentele ştiinţifice, unele indicaţii privind punerea în practică şi normele de evaluare a rezultatelor. Este sursa informaţională principală pentru aprecierea încrederii pe care trebuie să o avem într-un test. Manualul trebuie să dea indicaţii detaliate despre procedeele de evaluare şi despre modul în care se redactează

concluziile aplicării unui test. Între exigenţele ideale şi realitatea practică există o anumită distanţă, astfel încât calitatea prezentării poate lăsa de dorit, în unele cazuri, prin omisiunea unor factori care ar putea folosi utilizatorului şi prin ascunderea, omiterea unor aspecte dezavantajoase ale testului. L.J. Cronbach afirmă că testele elaborate înainte de 1945 prezentau foarte rar manuale complete şi nu erau conforme standardelor de atunci în domeniul examinării psihologice prin intermediul testelor (elaborate de asociaţiile profesionale) (*cf.* Cronbach, 1979, pp. 159-162). Redactarea unui bun manual este o chestiune cu atât mai dificilă cu cât s-au adunat mai multe rezultate despre test. Manualul trebuie să fie alcătuit în aşa fel încât să fie foarte clar pentru profesionişti şi să ducă la recunoaşterea de către nonprofesionişti a faptului că ei nu posedă o calificare în domeniu. Un important ajutor în redactarea manualelor îl constituie recomandările tehnice editate de asociaţiile profesionale internaţionale şi naţionale (cele mai cunoscute sunt cele editate de APA). Aceste recomandări tehnice arată ce trebuie să conţină un manual. Curentul de perfecţionare a calităţilor testului psihologic şi a manualelor aferente a cunoscut prin O.K. Buros o conturare clară începând cu 1934, autorul emiţând opinii critice despre testele cuprinse în *Mental Measurement Yearbook*. Realizările în această direcţie sunt deosebite în 1950, 1953 şi 1959, fiecare test fiind cercetat de mai mulţi specialişti, selectaţi pe baza cunoştinţelor lor tehnice şi a experienţei lor practice. Aceştia explicau în reviste de specialitate, precum *Educational and Psychological Measurement* şi *Journal of Consulting Psychology*, de ce sunt mai recomandabile unele teste şi supuneau atenţiei utilizatorilor chestiunile problematice din manualele testelor;
   - materialul testului, constând din diferite tipuri de aparate pentru evidenţierea stimulilor, caietul testului, foi de răspuns, materiale pentru prelucrare şi pentru manipulare. Este partea care asigură stimularea comportamentelor de răspuns;
   - auxiliare de evaluare, precum grilele de corectură (şabloane, folii cu orificii) şi aparate de măsură;

b) *componenta dinamică*, de realizare a unui test, se divide metodologic şi tehnologic, după cum urmează:
   - consemnul testului: instruieşte persoana calificată profesional, care aplică testul, cu privire la condiţiile, realizarea, evaluarea testelor şi la ceea ce li se pretinde subiecţilor, în timpul examinării cu ajutorul unui test. Informaţii de acest gen pot fi cuprinse uneori în manualul testului;
   - partea de realizare poate să se exprime într-o reacţie motorie (de exemplu, reacţie senzorial-motorie sau verbală la stimuli simpli ori complecşi, *tapping*), într-o percepţie senzorială (de exemplu, tahistoscop), într-o activitate de rutină mintală sau fizică (a înşira bile, a calcula după un algoritm, cum este cazul la testul Kraepelin), în rezolvarea unei probleme concrete sau abstracte (a construi pompa Schulze, a monta instalaţia denumită *Winkeltrieb*, a rezolva sarcini de gândire) sau în răspunsul la întrebări privind comportamentul, trăirile ori dispoziţiile. Sarcinile din cadrul unui test pot să aibă un conţinut simplu sau complex, dificultăţile lor putând fi aproximativ egale sau foarte diferite;
   - evaluarea poate să se realizeze intuitiv, regulat, schematic sau automatic şi constă în aceea că rezultatele testului sunt exprimate în unul sau mai multe punctaje.

Componentele testului suportă condensări informaţionale, o dată cu procesul de computerizare a testului psihologic. Practic, informaţiile părţii statice pot fi condensate pe o dischetă şi accesate prin intermediul diferitelor tipuri de computere. După cum am amintit deja, testele computerizate au rezervată o secţiune aparte în lucrarea noastră.

3. Clasificarea testelor psihologice

Clasificarea este o operaţie prin care obiectele dintr-o mulţime dată (sau un univers) sunt împărţite în clase, în funcţie de asemănările şi deosebirile dintre ele. O clasificare a testelor are, cel mult, forţa informaţională a unei scale ordinale, dar de cele mai multe ori a scalei nominale, în care se pot opera grupări şi rafinări. Compararea obiectelor unui univers se face dintr-un anumit unghi de vedere, numit *criteriu*. Clasificarea poate fi numită *naturală*, dacă descrie clasele aşa cum sunt în realitate, sau *artificială*, în cazul în care clasele sunt formate după anumite criterii convenţionale, în funcţie de utilitate. De asemenea, clasificarea poate fi *teoretică* (regăsindu-se aici atât clasele reale, cât şi cele posibile) sau *empirică* (în care avem doar clasele reale). Clasificarea poate fi făcută după unul sau mai multe criterii. În funcţie de natura obiectelor, poate fi *exactă* sau *inexactă*. O clasificare exactă presupune că pentru orice obiect din universul supus clasificării este determinată exact clasa căreia îi aparţine. Pentru cazul testelor psihologice, putem spune că este vorba despre o clasificare artificială, empirică şi inexactă.

În mod tradiţional, în clasificare se porneşte de la tipul ideal, care are la bază supoziţii referitoare la satisfacerea terţului exclus pentru orice obiect din universul supus clasificării în raport cu proprietăţile invocate, până la o prezentare completă, fiecare obiect din universul de clasificare făcând parte din una dintre clasele indicate; în ansamblu, clasele se exclud între ele, suma claselor identificându-se cu universul. În realitate, aceste cerinţe nu sunt realizate decât în parte, deoarece în practica de clasificare ne confruntăm adesea cu *mulţimi vagi*, în unele cazuri cu mulţimi infinite sau mulţimi nedefinite. Universul testelor psihologice nu realizează toate exigenţele ideale ale clasificării, deoarece se operează cu mulţimi vagi (unele clasificări includ şi testele de personalitate subiective, altele nu), care nu sunt finite (clasificările actuale cuprind mai multe clase de teste decât cuprindeau acum 50 de ani, numărul obiectelor universului de clasificare fiind în creştere cu o rată imprevizibilă).

Din analiza clasificării testelor psihologice, realizată de către diverşi autori, putem conchide că nu există o unitate de opinii, dar între majoritatea criteriilor de clasificare utilizate există o concordanţă apreciabilă. În privinţa clasificării testelor, s-au pronunţat personalităţi de prim rang ale psihologiei aplicate. Astfel, enumerăm în această privinţă autori ca H.J. Eysenck, (1953), M. Irle (1956), G. Lienert (1969), L.J. Cronbach (1970), P. Horst (1971), R.L. Thorndike (1971), R. Meili şi E. Steingruber (1978), Anna Anastasi, Rolf Brickenkamp (1997), D. Laveault şi J. Grégoire (1994)[1]. În 1973, D. Speck şi R. Speck introduc două grupe de criterii de clasificare, preluate de majoritatea celor ce se ocupă de clasificarea testelor. Autorii germani elaborează o diviziune utilitar-pragmatică a testelor, din punctul de vedere al conţinutului şi din

1. Anii trecuţi între paranteze sunt cei ai întocmirii unor sisteme de clasificare de către autorii respectivi. Precizările sunt realizate în limita bibliografiei parcurse.

punctul de vedere al formei. Cea mai mare concordanţă în sistemele de clasificare se află în grupa celor ce adoptă punctul de vedere formal. Aspectele clasificării formale sunt caracteristicile construcţiei testului, materialele pe care le utilizează testul, felul realizării comportamentului relevant din punct de vedere diagnostic, tipurile de reacţie care sunt pretinse şi modul de realizare a evaluării. Cronbach, Eysenck, Michel, Lienert, Ingekamp şi Irle fac referiri asupra criteriilor de clasificare formale. Perspectiva poate să fie importantă mai ales pentru practicieni, deoarece pentru aceştia este esenţial să ştie dacă un test poate fi aplicat individual sau în grup, dacă face apel la exprimarea verbală în cursul efectuării, dacă pretinde reacţii motorii în răspuns. Clasificarea care adoptă respectivul punct de vedere prezintă un dezavantaj clar, deoarece nu face nici o referire la scopul unui procedeu de examinare, deci nu precizează nimic despre conţinutul unui test, despre ceea ce trebuie să măsoare un test. Este, de asemenea, dificil să se identifice domeniile de aplicaţie ale unui test. De exemplu, testele de inteligenţă se pot încadra în orice categorie a unei clasificări care se bazează pe punctul de vedere formal.

Enumerăm în continuare cele mai importante criterii de clasificare formală:

- După gradul de standardizare şi cuantificare, există teste *psihometrice* (evaluarea se produce numeric şi prin norme) şi *nepsihometrice* (care fac apel în emiterea concluziilor la virtuţile clinice de exprimare). Acesta este un punct de vedere întâlnit în bibliografia de specialitate, nefiind însă împărtăşit în totalitate de autorul prezentei lucrări, care a definit anterior testul ca un procedeu de investigaţie standardizat. Nu înseamnă că trebuie să ignorăm alte opinii exprimate de autori cu experienţă în domeniul teoriei şi practicii testelor psihologice. Împărţirea în teste standardizate (etalonate) şi nestandardizate (informale) apare, de exemplu, într-un tratat al cărui autor este Gustav A. Lienert. În concepţia acestuia, testele informale sunt destinate uzului personal al psihologilor şi al profesorilor, care îşi propun, de exemplu, să claseze doar membrii unei singure clase şcolare cu care lucrează frecvent. Rezultatele testelor standardizate capătă semnificaţie cantitativă prin raportare la rezultatele unui grup de referinţă, numit şi grup normativ, după ce au fost create şi anumite sisteme de etalonare.
- După posibilitatea de a examina simultan mai mulţi subiecţi, distingem între *teste individuale* şi *teste colective* (sau de grup). Testul individual are dezavantajul de a aloca un timp de aplicare crescut pentru un singur subiect, dar are avantajul că oferă posibilitatea de a observa mai bine reacţiile particulare în timpul examinării. Testul colectiv are ca principal avantaj economia temporală a examinării şi ca dezavantaj observarea insuficientă a subiectului. Clasele rezultate din operarea cu acest criteriu de diviziune interferează cu clasele rezultate din aplicarea altor criterii de clasificare şi sunt, de asemenea, interschimbabile în interiorul aceluiaşi criteriu. Astfel, majoritatea testelor colective sunt teste creion-hârtie, iar majoritatea testelor individuale sunt teste aparative sau de prelucrare a materialelor. O tendinţă demnă de luat în seamă este aceea a transformării prin computerizare a testelor individuale în teste colective. Astfel, testele care presupun evidenţierea timpului de reacţie, simplu sau complex, pot fi efectuate colectiv prin deservirea de către acelaşi computer a mai multe instalaţii de examinare.
- Din punctul de vedere al mediului informaţional al testului, se poate vorbi despre *teste creion-hârtie*, *teste de reacţie verbală*, *teste de manipulare şi prelucrare a materialelor*, *teste-imagini*, teste care apar *pe display-ul computerului*, *teste aparative*.

Clasele constituite prin activarea acestui criteriu sunt într-o interschimbare accentuată, ceea ce face ca anumite clase să fie „golite", „aspirate" de altele. Este cazul testelor creion-hârtie, care sunt transformate într-o proporţie apreciabilă în teste computerizate.

- În funcţie de utilizarea formei verbale de exprimare în producerea răspunsului la situaţia problematică, deosebim între *teste verbale* (care depind, de cele mai multe ori, de nivelul de cultură al subiectului) şi *teste nonverbale* (care nu depind de nivelul de cultură, fiind create special pentru a contracara influenţa acestui factor). De-a lungul istoriei exploatării diagnostice a testelor psihologice, li s-a reproşat adeseori testelor că favorizează subiecţii care provin din straturi sociale cultivate. Forma nonverbală de exprimare este parţială sau totală, în primul caz aparţinându-i doar subiectului, căruia examinatorul i se poate adresa verbal pentru a-i da explicaţii cu privire la ceea ce are de făcut, sau atât subiectului, cât şi examinatorului, mai ales în examinarea surdomuţilor. Pentru exprimarea răspunsurilor la această categorie de teste se poate folosi întreaga gamă a comunicării paraverbale.
- În funcţie de numărul de dimensiuni cuprinse în destinaţia psihodiagnostică a testului, deosebim între *testele unidimensionale* (de exemplu MMQ, alcătuit de H.J. Eysenck, şi KLT, alcătuit de H. Düker) şi *testele multidimensionale* (IST, al lui R. Amthauer, şi PIT, al lui E. Mittennecker). Acest criteriu se poate reformula în clasificarea *testelor singulare* (care posedă o singură componentă) şi a *bateriilor de teste*. Testele singulare[1] nu se confundă cu testele individuale, ci pot doar interfera. Atributul de „singular" se referă la numărul de componente ale testului, pe când cel de „individual" arată că se poate aplica unui singur individ. Testul singular poate fi aplicat individual sau colectiv, iar testul individual poate fi unidimensional sau multidimensional. Exigenţele în privinţa realizării unidimensionalităţii unui instrument de investigaţie psihodiagnostică sunt foarte ridicate, deoarece faptul psihic se află, prin esenţa sa, în interdependenţă.
- Dacă luăm în considerare gradul de structurare a configuraţiei reacţiilor care provoacă comportamentul analizat, facem distincţie între *testele înalt structurate* şi *testele slab structurate*. În testele slab structurate, stimularea nu este dată sub forma unei solicitări ce presupune o formă de răspuns dihotomică sau polihotomică, ci sub o formă care impune un efort din partea subiectului de structurare a răspunsului.
- În funcţie de felul comportamentului provocat sau al comportamentului indus prin situaţia de testare, avem teste care cer un comportament maximal, care presupun mobilizarea subiectului pentru atingerea unui nivel de randament cât mai înalt şi teste care provoacă un comportament tipic, cum este cazul testelor de personalitate, unde în răspuns este vorba despre o opţiune bazată pe sistemul moral-valoric al examinatului sau pe gusturile, înclinaţiile, preferinţele acestuia. Răspunsurile ce provoacă un comportament tipic nu pot fi considerate „bune" sau „rele".
- În funcţie de limita de timp pentru efectuarea testului, deosebim, doar în cadrul testelor ce provoacă un comportament maximal, *teste de rapiditate* (*speed tests*) şi *teste de nivel* (*power tests*), însă în majoritatea situaţiilor practice avem de-a face cu

1. Denumirea reprezintă traducerea termenului german „*Einzeltest*", întâlnit la Rolf Brickenkamp (coord.), *Handbuch psychologischer und pädagogischer Tests*, Hogrefe, Verlag für Psychologie, Göttingen, Berna, Toronto, Seattle, 1997, p. XV.

o îmbinare a celor două clase. Testele de rapiditate cuprind sarcini (itemi) de dificultate medie şi uşoară într-un asemenea număr încât nici un subiect nu poate să răspundă la toate într-un timp prestabilit. Un exemplu pentru testele de rapiditate este testul Bourdon-Amfimov, unul dintre cele mai vechi (datează din 1895), în care i se cere subiectului să bareze literele O şi C în cadrul unor rânduri tipărite ce conţin diverse litere. Se pune accentul pe aprecierea randamentului, pe rapiditatea în rezolvarea sarcinilor. Testele de nivel (numite în literatura de specialitate de limbă engleză „*power test*" sau „*level test*") cuprind sarcini care cresc continuu gradul de dificultate, încât ultimele sarcini abia sunt rezolvate, în ciuda limitei de timp generoase. Este un nonsens să se vorbească despre teste fără limită de timp, având în vedere că fiinţa umană este finită în timp.

- O separare frecventă în spaţiul cultural anglo-saxon şi operată foarte rar în cel german se face între *testele de aptitudini* („*aptitude tests*") şi *testele de achiziţie* („*achievement tests*"), cele din urmă urmărind diagnosticarea cunoştinţelor şi abilităţilor dobândite prin exerciţiu şi şcolarizare.
- Dihotomia *teste subiective – teste obiective* pentru măsurarea personalităţii se practică în vederea evidenţierii posibilităţii subiective de deformare a rezultatelor de către examinator, şi nu de către subiect[1].
- O serie de autori, dintre care îi enumerăm pe P. Horst (1971), J.P. Guilford (1954), R.B. Cattell (1962) şi E. Todt (1968), facem deosebirea între *testele normative* şi *testele ipsative*. În cadrul acestei distincţii, referirea se face, implicit, la două feluri de măsurare: normativă şi ipsativă. Am amintit deja că, în optica normativă, aprecierea valorică a unui rezultat individual se face prin raportare la rezultatele unui grup, numit grup de referinţă. În cazul unui instrument de măsură ipsativ, valorile de punctaj pentru o anumită caracteristică supusă măsurării nu sunt comparabile de la un subiect la altul, ci au semnificaţie numai în relaţie cu un singur subiect; este vorba, aşadar, în această situaţie, despre un univers de apreciere strict individualizat.
- Dacă luăm drept criteriu libertatea de reacţie, vom găsi *teste cu dispozitiv de răspuns creativ* (sau deschis) şi *teste cu dispozitiv de răspuns legat* (sau închis). În primul caz subiectul are libertatea producerii răspunsului, iar în al doilea trebuie să aleagă dintr-o serie de răspunsuri oferite.
- O separare similară celei anterioare diferenţiază *testele formale* de cele *informale*. În testele formale, forma de interpretare a răspunsului este strict condiţionată de o terminologie prestabilită. Testele informale lasă un anumit grad de libertate în această privinţă. Revine, într-o astfel de împărţire dihotomică, problema standardizării instrumentelor de măsură, deoarece standardizarea reprezintă o exigenţă şi pentru partea de interpretare a răspunsurilor.
- Un criteriu formal de clasificare, apropiat de cel de mai sus, îl reprezintă interpretarea: se deosebesc astfel *testele directe* (sau psihometrice) şi *testele indirecte* (sau proiective). Cele proiective pot fi, la rândul lor, realiste (de exemplu, TAT) şi simbolice (de exemplu, Rorschach). Acest criteriu interferează cu alte criterii. O parte dintre clasificatorii testelor psihologice consideră testele proiective ca fiind tehnici proiective, şi nu teste.

1. Distincţie realizată în *ibidem*, p. XVI.

➢ Deosebirea dintre *testele psihometrice clasice* şi *testele probabiliste* se face în funcţie de structurile sintactice şi semantice ale modelului aflat la baza construcţiei lor. Vom prezenta mai târziu câteva consideraţii speciale referitoare la aceste clase de teste.

Specialistul care se interesează de un test în scopul de a-l folosi doreşte în primul rând să ştie ce variabile sunt măsurate şi, după aceea, în ce modalităţi formale se realizează acest lucru cel mai bine. Noi am tratat cu deosebire criteriile formale, deoarece în privinţa lor există un consens puternic între specialiştii din domeniu care au efectuat clasificări.

Aspectele de conţinut ale clasificării, chiar cu dezacordurile care se manifestă, sunt indispensabile pentru orientarea practică a psihologului şi psihopedagogului. Prin aspectele de conţinut s-ar putea înţelege domeniile de aplicaţie a diverselor teste, dar, în acest caz, un test de inteligenţă s-ar putea plasa în aproape toate domeniile de aplicaţie. Cel mai adesea, clasificarea de conţinut a testelor psihologice răspunde la întrebarea: Ce caracteristică examinează, de fapt, respectivul procedeu de investigaţie psihodiagnostică?

Deşi este un teren mai diversificat, în câmpul clasificărilor de conţinut domină totuşi o clasificare larg acceptată: este vorba despre separarea în *teste de aptitudini* (teste de inteligenţă şi teste de randament), pe de o parte, şi *teste de personalitate*, pe de altă parte. Este adevărat că dihotomia aptitudine (respectiv capacitate) – personalitate este destul de greu de acceptat. Practic, noţiunea de *personalitate* implică şi aspecte de randament sau de capacitate. Drept urmare, Cattell şi Wurburton subsumează sub genul proxim „teste de personalitate" procedee pentru diagnosticarea atât a aptitudinilor, cât şi a aspectelor esenţiale dinamic motivaţionale şi a modalităţilor temperamentale. În privinţa conţinutului noţiunii de *teste de randament*, există un mai mare consens decât asupra noţiunii de *test de personalitate*. O mare răspândire au clasificările din punctul de vedere al conţinutului, avându-i ca autori pe Cronbach, Anastasi, Irle, Brickenkamp şi Lienert. Prezentăm în cele ce urmează respectivele clasificări de conţinut, specificând în unele cazuri şi denumirile originale, deoarece traducerea nu poate cuprinde anumite conotaţii importante).

- *Anna Anastasi*
  1. Teste de dezvoltare intelectuală generală
  2. Teste de aptitudini (capacităţi) individuale – *Test of separate abilities*
     - 2.1. Măsurarea aptitudinilor multiple
     - 2.2. Teste de aptitudini speciale
     - 2.3. Teste educaţionale
     - 2.4. Teste ocupaţionale
  3. Teste de personalitate
     - 3.1. Inventare de autoraportare
     - 3.2. Măsurare a intereselor şi atitudinilor
     - 3.3. Tehnici proiective
     - 3.4. Alte tehnici pentru evaluarea personalităţii

- *Cronbach*
  1. Teste de capacitate (teste de performanţă maximă) – *Tests of ability* (*tests of maximum performance*)
     - 1.1. Măsurarea capacităţii mintale generale (*measure of general mental ability*)
     - 1.2. Măsurarea capacităţilor speciale (*measure of special abilities*)
     - 1.3. Teste de aptitudini
     - 1.4. Teste de achiziţie (*achievement tests*)
     - 1.5. Teste de expertiză (*proficiency tests*)
  2. Teste de performanţe tipice
     - 2.1. Observaţii comportamentale
     - 2.2. Procedee de autoraportare

Din cauza unor ambiguităţi de traducere a termenilor utilizaţi, denumirile claselor clasificării sunt folosite în varianta engleză de o parte dintre cei care operează cu clasificări ale testelor. Prin *„general mental ability"* se înţelege (frecvent, dar nu exclusiv) capacitatea generală în care importante sunt activităţile de gândire şi de învăţare. În cazul testelor pentru examinarea capacităţilor speciale şi a aptitudinilor, se poate deosebi între testele orientate spre diagnosticarea funcţiilor sau capacităţilor psihice elementare şi bateriile de teste, desemnate în spaţiul lingvistic anglo-saxon cu denumirea *„multiple aptitude batteries"*, ele trebuind să servească examinării unor capacităţi empirice care pot fi delimitate una de alta. Doi autori americani, L.L. Thurstone şi Thelma G. Thurstone, au impus denumirea de *„tests of primary mental abilities"*, iniţiind astfel tendinţa, aflată în permanent progres în SUA, de a dezvolta baterii de teste pentru capacităţi multiple, care se bazează frecvent pe analiza factorială. Aceste teste sunt folosite atât în domenii înguste (examinarea aptitudinii pentru anumite profesii, teste de selecţie şcolară etc.), cât şi extinse.

- *Brickenkamp*
  1. Teste de randament
     - 1.1. Teste de dezvoltare
     - 1.2. Teste de inteligenţă
     - 1.3. Teste generale de randament
     - 1.4. Teste şcolare
     - 1.5. Teste pentru examinarea funcţiilor generale şi teste de aptitudini
  2. Teste psihometrice de personalitate
     - 2.1. Teste de structură a personalităţii
     - 2.2. Teste de dispoziţii şi interese
     - 2.3. Teste clinice
  3. Procedee de evaluare a personalităţii
     - 3.1. Procedee de interpretare a formelor
     - 3.2. Procedee verbal-tematice
     - 3.3. Procedee grafice şi de structurare

- În 1956, *Irle* întocmeşte o clasificare în patru grupe, respectând şi standardele americane[1]:
  1. Teste generale de inteligenţă
  2. Teste de capacitate/teste de aptitudine
  3. Teste de cunoştinţe/teste de randament
  4. Teste de personalitate

- În 1961, *Gustav Lienert* întocmeşte o clasificare cu mare răspândire:
  1. Teste de inteligenţă
     1.1. Teste de inteligenţă generală
     1.2. Teste speciale de inteligenţă
  2. Teste de randament
     2.1. Teste de randament motrice
     2.2. Teste de randament senzorial
     2.3. Teste de randament psihic
  3. Teste de personalitate
     3.1. Teste de caracteristici
     3.2. Teste de interese
     3.3. Teste de dispoziţii
     3.4. Teste de caracter
     3.5. Teste pentru stabilirea tipului de personalitate.

Am prezentat doar principalele sisteme de clasificare, care nu epuizează spectrul acestui domeniu. Punem punct aici încercărilor de clasificare a testelor psihologice sub ameninţarea *briciului lui Ockham*, denumire a unui principiu atribuit lui W. Ockham şi destinat să combată distincţiile sau multiplicările inutile ale entităţilor în filosofia medievală, principiu formulat astfel: *Entia non sunt multiplicanda praeter necessitatem* (entităţile nu trebuie înmulţite peste necesităţi).

În continuare, se aduc o serie de clarificări pentru anumite clase şi subclase ale sistemelor de clasificare din punctul de vedere al conţinutului, comune sau nu sistemelor amintite mai sus. Se consideră clarificate de la sine un număr de subdiviziuni ale sistemului de clasificare, cum ar fi, de exemplu, testele de randament, în calitate de categorie de teste în care se pune accentul pe cantitatea de sarcini rezolvate corect. Precizări adecvate asupra unor clase vom face şi în alte secţiuni ale lucrării de faţă.

Începem comentariile cu testele numite *de dezvoltare*. Este vorba despre acele instrumente de investigaţie care trebuie să evalueze nivelul de dezvoltare a unui subiect. În 1964, G. Reinert face distincţia între *testele primare de dezvoltare* şi *testele secundare*

1. *Apud* Lothar Michel, „Allgemeine Grundlagen psychometrischer tests", in R. Heiss (coord.), *Psychologische Diagnostik*, Dr. C. Hogrefe, Verlag für Psychologie, Göttingen, 1964.

de dezvoltare. Cele primare definesc şi, respectiv, normează cursul dezvoltării şi au ca obiectiv evaluarea nivelului de dezvoltare al comportamentului, pe când în cazul celor secundare obiectivul propriu-zis se situează în afara sistemului de referinţă al diagnozei dezvoltării. Astfel, testele generale de randament sunt normate pe un larg domeniu de vârste şi permit evidenţierea unei anumite tendinţe de dezvoltare generală. Denumirea folosită cel mai frecvent în legătură cu diagnosticarea nivelului de dezvoltare este cea de scală (sau scară) de dezvoltare. Este necesar să facem anumite precizări în privinţa noţiunii de *scală* şi a diferenţierilor dintre scală şi test. Există tendinţa de a considera sinonimi termenii de *test* şi de *scală*. Se folosesc curent expresiile *scala Binet-Simon sau testul Binet-Simon*, *scala Terman-Merrill sau testul Terman-Merrill*[1]. De asemenea, există tendinţa de a considera scala o grupare de itemi referitori la un acelaşi aspect, deci de a considera termenul de *scală* rezervat construcţiilor psihodiagnostice unidimensionale. În cazul chestionarelor de personalitate, această tendinţă este cea mai justificată. Dar există şi scale multidimensionale: WISC-ul lui Wechsler este o scală pentru diagnosticarea inteligenţei la copil, având mai multe componente psihodiagnostice diferenţiate. Instrumentul pe ansamblu poartă denumirea de *scală*, iar componentele sale sunt denumite *teste* sau *subscale*. În bibliografia consultată, am întâlnit expresia „testele scalei Binet-Simon". Iată deci cum o scală cuprinde mai multe teste. Cu toate acestea, în concepţia noastră, noţiunea de *scală* nu este supraordonată celei de *test*.

Scala, ca instrument de investigare psihodiagnostică, este o formă particulară a scalei de măsură în general, care îşi propune să pună în evidenţă ansamblul modalităţilor sau valorilor observabile ale unei variabile[2]. Prin scală de măsurare psihologică se înţelege un instrument de măsură utilizat în psihologia aplicată, care serveşte la ordonarea unor niveluri succesive de complexitate crescândă. Forţa informaţională a scalelor este foarte diferenţiată.

În domeniul psihometric, prin „scală" se înţelege un test sau o baterie de teste, care permite clasarea subiecţilor în diferite niveluri. Considerăm că nota definitorie a unei scale este capacitatea ei de a diferenţia cel puţin ordinal subiecţii, de a-i plasa în funcţie de un gradient al său, în funcţie de rezultatul obţinut. Obiectele evaluării pot fi niveluri de dezvoltare intelectuală (unde se operează în mod deosebit cu noţiunile de vârstă mintală şi IQ), niveluri de dezvoltare motrice, precum şi niveluri de dezvoltare a personalităţii. În psihologia diferenţială, scalele sunt probe constituite din itemi ordonaţi după gradul de dificultate, permiţând exprimarea nivelurilor de dezvoltare progresivă. Cele mai cunoscute sunt: scala Binet-Simon (dezvoltarea inteligenţei), scala Alexander (inteligenţă practică, nonverbală), scala Brunet-Lézine (dezvoltare psihomotorie a primei copilării sau maturitate mintală), scala Columbia (scală colectivă de nivel intelectual), scala Grace-Arthur (inteligenţă nonverbală), scala Taylor (anxietate manifestă) etc.

Continuăm comentariile cu *testele de aptitudini*. La autorii englezi, care întocmesc clasificări ale testelor psihologice, apare o ambiguitate de interpretare în privinţa termenului *„ability"*, care se traduce prin „aptitudine", dar şi prin „capacitate". Aceeaşi

1. Ultima expresie este utilizată de Maurice Reuchlin, în *Grand dictionnaire de la psychologie*, Larousse, Paris, 1992, p. 254, în secţiunea din dicţionar rezervată precizării noţiunilor de *scală* şi *scală de măsură*.
2. Aspectul general al scalei de măsură este tratat detaliat în lucrarea lui Valentin Clocotici şi Aurel Stan, *Statistică aplicată în psihologie*, Editura Polirom, Iaşi, 2000.

ambiguitate există la traducerea termenului german „*Fahigkeitstest*". Facem încă o dată precizarea că aptitudinea este o caracteristică subiacentă, indirect observabilă. Pe lângă acest sens general, trebuie luate în considerare trei conotații:

- În primul rând, trebuie evidențiată concepția lui L.L. Thurstone (în lucrarea sa, *Primary Mental Abilities*, apărută în 1938) care consideră aptitudinile ca fiind caracteristici elementare fundamentale. În acest sens, testele de aptitudini furnizează o serie de caracteristici analitice, evidențiate în așa-numitele *profiluri psihologice*, care se opun scalelor de inteligență de tipul Binet-Simon, exprimate în IQ sau vârstă mintală. Această concepție a fost susținută într-o serie de cercetări de analiză factorială, încă de la apariția modelului considerat. R.B. Cattell depășește opoziția pe care se fondează această concepție și propune o structură de aptitudini umane în formă de arbore.
- În al doilea rând, aptitudinile pot fi considerate caracteristici cognitive responsabile de achiziția și de tratarea informației. Ele se deosebesc de caracteristicile motivaționale și temperamentale, asigurând anumite scopuri și conferind un anumit stil de funcționare mecanismelor cognitive. Această distincție se reflectă în separarea probele de aptitudini de probele de personalitate. Și ea suferă pe parcurs o anumită relativizare, printr-o „personalizare a teoriilor cognitive" (sintagmă aparținându-i lui Maurice Huteau).
- În al treilea rând, aptitudinile pot fi definite prin caracterul lor constituțional determinat de factorii genetici. H. Piéron este considerat principalul reprezentant al acestei optici științifice, în care testele de aptitudini se disting în special de testele de cunoștințe achiziționate, rezultatele lor nedepinzând de nici o cunoștință specifică prealabilă. Așadar, în această concepție, a existat tendința de a construi, pe cât posibil, teste în care rezolvarea să nu depindă de nivelul de cultură al subiectului. Ultima distincție ne amintește de disputa cercetătorilor privind influența factorilor genetici și a celor ambientali în determinarea variabilității interindividuale. Radicalismul unei astfel de opoziții este serios diluat de teoriile care pun accentul pe interacțiunea celor două tipuri de factori.

Considerațiile anterioare constituie însă o tratare minimală a problemei aptitudinilor umane.

Trebuie să remarcăm în plus că *testele de inteligență* sunt cele mai perfecționate realizări ale testelor de aptitudini. Este necesar să amintim că, deocamdată, nu există o definiție unitară a inteligenței, consecință firească a abordărilor teoretice din unghiuri diferite ale unui proces psihic deosebit de complex. Revoluția cognitivă din cercetarea psihologică introduce un nou aparat conceptual (noi optici de abordare a fenomenelor, de care trebuie să țină seama și autorii unor teste), mult extins față de ceea ce înțeleg, în mod obișnuit, prin inteligență unii autori de teste. Este, prin urmare, foarte important să acordăm o atenție deosebită precizării termenilor utilizați în cadrul manualului testului.

Menționăm că *testele școlare* își propun drept obiectiv diagnosticarea capacităților și aptitudinilor elevilor. Considerăm că denumirea de *teste pedagogice* ar fi mai potrivită pentru o parte dintre ele. Testele școlare cuprind teste de calcul și de lectură, teste de ortografie, teste speciale de aptitudini școlare, teste de învățare școlară. Unele nu au o referință valorică normativă a rezultatelor, raportându-se doar la clasa de elevi la care se aplică.

Se impune o precizare în privința termenului de *inventar de personalitate*, considerat sinonim cu cel de *chestionar de personalitate*, ambele fiind instrumente de investigație

constituite sub formă de scale referitoare la ansamblul personalității sau la diferite componente ale acesteia. Am amintit deja că unii autori, care au întocmit clasificări ale testelor psihologice, le contestă acestora calitatea de teste, datorită naturii subiective a răspunsurilor ce se obțin după efectuarea lor. A considera termenii de *inventar de personalitate* și *chestionar de personalitate* sinonimi cu cel de *test de personalitate* ar însemna să interpretăm forțat. În unele cazuri, pentru a stabili mai clar diferențierea, se folosește sintagma *teste obiective de personalitate*, conținutul răspunsurilor transpunând un „stil al personalității", cu mare constanță de manifestare, cum se întâmplă, de exemplu, în testul miochinetic al lui Mira y Lopez.

Ceea ce se desemnează de obicei prin noțiunea de *tehnici proiective* în câmpul instrumentelor de investigație psihodiagnostică reprezintă categoria cel mai frecvent exclusă de la statutul de test psihologic, datorită dificultății deosebite de a asigura criteriile de validitate (mai ales validitatea) și datorită dozei mari de subiectivitate în interpretare. „Se poate discuta, pe bună dreptate, dacă termenul de test, luat în sens riguros psihometric, se aplică într-adevăr la astfel de probe, în care standardizarea și etalonarea lasă adesea de dorit, în care sensibilitatea, fidelitatea și validitatea sunt dificil de estimat" (Anzieu și Chabert, 1997, p. 14). Un reproș frecvent exprimat, malițios și dovedind o doză de ignoranță în privința acestor probe, este acela că fenomenul de proiecție se produce mai mult la examinator decât la candidat. Cele mai frecvente referiri la metodele de investigație proiectivă se fac prin intermediul creațiilor din domeniu realizate de Rorschach, Zulliger, Szondi, Lüscher, Rosenzweig, Murray. Denumirea „metodă proiectivă" îi aparține lui L.K. Frank și a fost folosită pentru prima dată într-un articol din revista științifică americană *Journal of Psychology*, în 1939. Termenul a fost conceput pentru a pune în evidență legătura dintre trei probe psihologice: testul de asociație al lui Carl Gustav Jung, elaborat în 1904, testul petelor de cerneală al lui Hermann Rorschach, elaborat în 1920, și *Thematic Apperception Test*, elaborat de Henry Murray. Tehnicile proiective se disting clar de testele de aptitudini prin ambiguitatea materialului prezentat subiectului și prin libertatea de răspuns care îi este lăsată acestuia. Baza lor teoretică, în centrul căreia stă noțiunea de *proiecție*, este asigurată, în special, de psihologia dinamică și de psihanaliză. Termenul de proiecție are mai multe sensuri, prezentate în continuare (*apud* Anzieu și Chabert, 1997, p. 16):

- Un prim sens denotă o acțiune fizică de aruncare, de lansare a unui proiectil. Trecând pe plan psihic, Sigmund Freud consideră *proiecția* ca fiind o acțiune psihică specifică stării de paranoia și care constă în expulzarea din inconștient a unor sentimente blamabile pentru a le atribui altuia.
- Al doilea sens este unul matematic, impus din secolul al XVII-lea, o dată cu apariția geometriei proiective. Prin *proiecție* facem să corespundă unui punct (sau unui ansamblu de puncte) din spațiu un punct (sau un ansamblu de puncte) al unei drepte ori al unei suprafețe. Transpunând noțiunea matematică în planul nostru de interes științific, testele proiective aduc subiectul în situația de a produce un protocol de răspuns (proiecția) a cărui structură corespunde structurii personalității (starea proiectată), caracteristicile fundamentale ale personalității regăsindu-se în această proiecție.
- La originea celui de al treilea sens se găsește fenomenul optic pus în evidență la sfârșitul secolului al XIX-lea, în care, plecând de la un focar de raze sau radiații, se trimite o proiecție luminoasă pe o suprafață. Astfel, proiecția fixează o imagine pe un ecran, așa cum se întâmplă în cinematografie. Sensul este adoptat și popularizat de

către scriitorii sfârşitului de secol XIX şi, ulterior, în psihopatologie şi în psihologie. Un test proiectiv, precum o rază X, traversează interiorul personalităţii, fixează imaginea nucleului secret al acestuia pe un revelator (trecerea sau efectuarea testului) şi permite astfel o lectură facilă prin mărirea sau îngroşarea proiecţiei pe un ecran (interpretarea protocolului). Ceea ce era ascuns este pus în lumină, latentul devine manifest, interiorul este adus la suprafaţă, ceea ce este stabil în noi se dezvăluie.

În ciuda reproşurilor făcute, tehnicile proiective au rămas în practica psihodiagnostică, un domeniu rezervat specialiştilor din domeniul clinic, care au un deosebit simţ al nuanţei în observaţie şi un rafinament deosebit în interpretare.

4. Teoria clasică a testului psihologic

4.1. Precizări

Atributul de *clasic* nu are, în cazul teoriei testului, o semnificaţie depreciativă, de *depăşit*, *simplu*, *revolut* sau *conservator*, ci una valorizatoare, de *bine fundamentat*, *echilibrat*, *durabil*. Acest atribut este justificat prin aceea că principiile sale au fost stabilite de întemeietorii stilului sistematic şi organizat în cercetarea psihologică, adică de reprezentanţii şcolii psihometrice engleze. Prezenţa şcolii respective în crearea suportului statistic al cercetărilor experimentale în psihologie este permanentă.

Sistemul de axiome sintactice de bază, cu tot eşafodajul de formule şi prescripţii procedurale ce decurg din acesta, a fost creat de Ch. Spearman în cele trei lucrări deja amintite de noi (publicate între 1904 şi 1907) şi a rămas neschimbat până la începutul deceniului şase al secolului XX, când H. Gulliksen îl prezintă cuprinzător şi-l dezvoltă pe baza unor ipoteze formal logice. C.J. Kousropoulos (1962, 1964) şi M.R. Novick (1966) au putut să arate, ceva mai târziu, că modelul teoriei clasice a testului psihologic poate fi explicat prin ipoteze mai puţine şi mai slabe, adică mai uşor de îndeplinit. Critici accentuate îşi fac apariţia chiar în timpul cizelării sintaxei modelului clasic şi al evidenţierii problematicii relevanţei sale pentru procesul de măsurare psihologică, reflectându-se în modelul testului prin ipoteze pătrunzătoare şi, totodată, falsificabile. Ca orice realizare omenească, şi testul psihometric construit pe axiomele clasice este considerat perfectibil, cu atât mai mult cu cât rezultatele din practică impun perfecţionarea lui. Nu trebuie să uităm că fiecare teorie este un model simplificat al realităţii, care se ajustează mai mult sau mai puţin realităţii pe care încearcă să o descrie. În multe cazuri din cercetarea psihologică, există o puternică tentaţie de a adapta datele la modele, în loc de a adapta modelele la date.

Raţiunea de apariţie a teoriei testului psihologic este aceea de a oferi un cadru conceptual, în scopul de a aprecia valoarea rezultatelor obţinute prin intermediul unui instrument de măsură. Majoritatea testelor psihologice îşi prezintă rezultatul brut sub forma unor scoruri compozite[1]. Acestea sunt dobândite prin însumarea valorilor numerice ale itemilor, adică a componentelor informaţionale elementare, bazale, ale unui test. Ceea ce interesează în mod special este cunoaşterea modului în care varianţa scorului

1. O excepţie notabilă este cea a probei de asociere şi disociere manuală, construită de Lahy.

total al unui test este influenţată de varianţele diferiţilor itemi. În funcţie de mărimea varianţei scorurilor totale ale unui test, se poate face diferenţierea cantitativ-valorică a caracteristicilor persoanelor supuse examinării. O varianţă egală cu 0 reprezintă o contraindicaţie absolută pentru folosirea unui test. Teoria clasică a scorurilor nu tratează decât precizia valorii numerice. Chestiunile privitoare la calitatea şi pertinenţa acestei valori sunt tratate separat, prin intermediul studiului validităţii. De altfel, atunci când vorbim de teoria testului putem face distincţia clară în teorii *structural-funcţionale* ale testelor psihologice şi teorii *cognitiv-semantice*. Primele teorii se referă la structura şi modul de funcţionare ale unui test, nu neapărat psihologic, iar celelalte la conţinutul şi semnificaţia psihologică ale denumirii trăsăturii psihice care se presupune că va fi diagnosticată. Ceea ce se desemnează cu sintagma de *teoria clasică a testului psihologic* se poate numi o teorie funcţională, adică o teorie care încearcă să explice modul de funcţionare a unui test sub aspectul relevanţei cantitativ-informaţionale.

4.2. Postulatele teoriei clasice a testului psihologic

În semnificaţia sa generală, postulatul este o propoziţie *primă*, adică nederivată din alte propoziţii, care se prezintă sub forma regulilor de definiţie, axiomelor, metaaxiomelor, schemelor de definiţie, regulilor de interpretare şi regulilor de deducţie. Având în vedere faptul că postulatele se referă la sisteme, putem spune că postulatele sunt nişte exigenţe ideale de funcţionare a sistemelor. În cazul în care o realizare empirică a unui sistem teoretic nu satisface postulatele aflate la baza funcţionării sale, se practică procedura reformulării postulatelor, care devin mai „slabe", adică mai uşor de satisfăcut. Această „relaxare" a postulatelor pentru astfel de cazuri nu poate fi considerată o abdicare de la spiritul ştiinţific.

Fiecare lucrare care se consacră teoriei clasice a testului psihologic începe prin a prezenta postulatele de bază proprii acestei teorii. Diferite lucrări ştiinţifice consacrate postulatelor teoriei clasice a testului psihologic prezintă un număr variabil de postulate, obţinute în urma unor operaţii de condensare sau rafinare. Astfel, J. Guthke, H.R. Böttcher şi L. Sprung prezintă, într-o lucrare consacrată psihodiagnosticului (1991, p. 133), o grupare de cinci postulate ale teoriei clasice a testului, L. Michel şi W. Conrad prezintă (1982, pp. 17-19) un număr de nouă axiome, iar W. Kristof (1983, pp. 545-547) un număr de patru. Redăm, parafrazat, postulatele teoriei clasice a scorurilor, aşa cum au fost concentrate într-o lucrare recentă (Laveault şi Grégoire, 1997, pp. 132-134) asupra teoriei testului:

- *Postulatul 1*. Scorul (X) observat al unei persoane supuse examinării prin intermediul unui test psihologic este suma scorului *adevărat* (notat prin V, o constantă) şi o eroare de măsură asociată acestui scor (notată prin E, o variabilă aleatorie):

$$X = V + E$$

Deci X este, de asemenea, o variabilă aleatorie. Eroarea E poate avea valoarea 0, valori pozitive, precum şi valori negative, în aşa fel încât, după un număr infinit de aplicări la aceeaşi persoană ale aceluiaşi test, erorile tind către o distribuţie normală, care poate fi standardizată (media este egală cu 0, iar abaterea standard egală cu 1). Măsurarea nu este nevoie să fie efectuată cu acelaşi test, ci şi cu teste strict paralele, adică teste care, aplicate pe aceiaşi subiecţi, dau distribuţii cu aceeaşi medie şi aceeaşi

varianţă (asupra noţiunii de *test paralel* vom reveni în secţiunea consacrată fidelităţii). Folosind o exemplificare fictivă, situaţia reaplicărilor aceluiaşi test sau ale unor teste strict paralele la acelaşi individ s-ar prezenta astfel:

Tabelul 2

Componentele scorurilor observate

Test	X	V	E
1	60	55	5
2	45	55	– 10
3	53	55	– 2
4	59	55	4
...	....	....	–
n	62	55	7

- *Postulatul 2.* Valoarea aşteptată pentru scorul observat este scorul adevărat:

$$\varepsilon(X) = V$$

Această formulă arată că scorul adevărat obţinut de un subiect este speranţa matematică a scorurilor observate. Precizia unui scor observat creşte o dată cu numărul de aplicări ale testului. Dacă acelaşi test se aplică de un număr mare de ori la acelaşi subiect, media rezultatelor va tinde către scorul adevărat. Dispersia scorului observat X în jurul scorului adevărat V constituie eroarea de măsură pentru ansamblul acestor aplicări.

- *Postulatul 3.* Între eroarea de măsură şi scorul adevărat al unui subiect nu există corelaţie (sau există o corelaţie nulă):

$$\rho_{EV} = 0$$

Aceasta înseamnă că eroarea întâmplătoare de măsură nu va fi mai mare, dacă subiectul posedă un scor adevărat ridicat, sau mai slabă, dacă el posedă un scor adevărat slab. Este axioma de independenţă dintre obiectul măsurat şi eroarea de măsurare.

- *Postulatul 4.* Erorile la două teste diferite (E_1 şi E_2) nu sunt corelate între ele:

$$\rho_{E_1E_2} = 0$$

- *Postulatul 5.* Între eroarea de măsură la un test şi scorul adevărat la un alt test nu există corelaţie:

$$\rho_{E_1V_2} = 0$$

- *Postulatul 6.* Două teste X şi X' sunt paralele dacă şi numai dacă scorurile lor adevărate şi parametrii erorilor de măsură sunt egale:

$$V = V'$$

$$\sigma_E = \sigma_{E'}$$

Din postulatul 1, care stipulează că scorul observat este suma dintre un scor adevărat şi un scor de eroare aleatorie, rezultă că două teste paralele vor avea sensibil aceeaşi medie şi aceeaşi varianţă a scorurilor observate.

- *Postulatul 7.* Un test este τ-echivalent (*tau* echivalent) cu un alt test când scorurile lor adevărate diferă printr-o constantă adăugată k:

$$V_1 = V_2 + k$$

Astfel, dacă 4 subiecţi obţin notele 23, 45, 67, 55 la un test şi, respectiv, 34, 56, 78, 66 la un alt test, iar erorile sunt nule, ar rezulta o constantă k cu valoarea 11. Testele paralele satisfac exigenţa de τ-echivalenţă (pentru k = 0), pe când reciproca nu este adevărată.

Putem rezuma ansamblul celor 7 postulate ale teoriei clasice a testului cu aserţiunea că erorile aleatorii de măsură trebuie să fie independente în orice circumstanţă. Ca atare, condiţiile de examinare trebuie să fie pregătite în aşa fel încât să nu existe corelaţii între scorurile adevărate ale unui subiect şi eroarea de măsură la un test sau între erorile de măsură de la teste diferite. Fără îndeplinirea acestor condiţii minimale, scorurile observate devin dificil de interpretat. Teoria clasică a scorurilor ia în considerare o eroare strict întâmplătoare.

Pentru un grup de subiecţi examinaţi cu un acelaşi test, situaţia rezultatelor scorurilor se poate prezenta în felul următor (date fictive):

Tabelul 3

Descompunerea varianţei scorului observat

Subiect	1	2	3	4	5	6	7	8	9	10	11	12	Medie	Varianţă
Scor observat (X)	8	8	5	9	3	7	8	5	10	7	8	9	7,25	3,68
Scor adevărat (V)	9	9	5	8	5	8	6	5	8	7	8	9	7,25	2,35
Eroare (E)	–1	–1	0	1	–2	–1	2	0	2	0	0	0	0,00	1,33

Din tabel se observă că scorul observat este suma dintre scorul adevărat şi eroare: X = V + E. În cazul unui scor compozit, varianţa scorului total are următoarea formulă: $S_X^2 = S_V^2 + S_E^2 + 2COV_{VE}$. Dacă avem în vedere faptul că valoarea covarianţei dintre V şi E este egală cu 0 (postulatul 3), atunci formula se transformă astfel: $S_X^2 = S_V^2 + S_E^2$. În cazul nostru concret, 3,68 = 2,35 + 1,33.

Adăugarea unui mare număr de astfel de erori, atât pozitive, cât şi negative, nu produce deplasarea mediei rezultatelor observate faţă de media rezultatelor adevărate. Dar există şi erori sistematice al căror efect se manifestă constant şi, în acest caz, rezultatul observat este deplasat faţă de cel adevărat (erorile nu se mai anulează reciproc). Preocuparea pentru identificarea unor astfel de erori nu mai intră în sfera teoriei clasice a scorurilor, ci în cea a validării rezultatelor (de exemplu, astfel de erori apar în cazul în care o probă este prea uşoară sau prea dificilă). În acest caz, ecuaţia de la postulatul 1 se poate rescrie în felul următor:

$$X = V + e_s + e_a$$

Scorul observat este egal cu suma dintre un scor adevărat, o eroare aleatorie şi o eroare sistematică. Dacă un examinator insistă asupra unui aspect foarte facil al disciplinei

sale, toţi examinaţii vor lua note maxime, fiind vorba în acest caz despre o asemenea situaţie. Eroarea aleatorie va fi absolut derizorie în comparaţie cu cea sistematică.

Repetarea măsurătorilor pentru acelaşi subiect (măsurători intraindividuale) sau pentru subiecţii unui grup (măsurători interindividuale) duce la distribuţii diferite în amplitudine, după cum se vede în figura următoare:

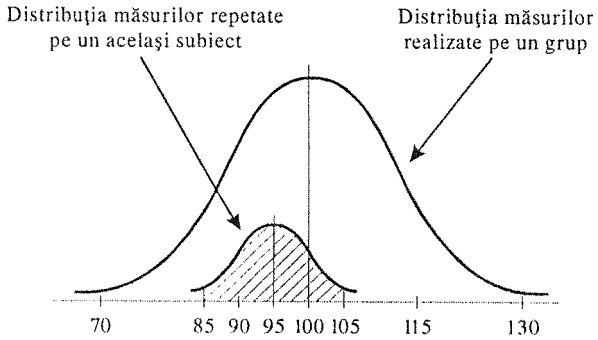


Figura 3. *Reprezentarea grafică a distribuţiilor rezultatelor obţinute prin repetarea măsurătorilor pe acelaşi subiect şi pe acelaşi grup*

Noţiunile de *validitate* şi de *fidelitate* se pot reprezenta în funcţie de proporţia varianţei scorurilor observate imputabilă varianţei pertinente (σ_v^2), varianţei nepertinente ($\sigma_{v_a}^2$) sau varianţei erorii ($\sigma_{e_a}^2$). Un constructor de test are drept scop maximizarea părţii scorului adevărat, care este pertinentă la ceea ce doreşte să măsoare, şi minimizarea erorii aleatorii de măsură. Pentru a reuşi acest lucru, este necesar ca varianţa scorului adevărat să ocupe o mare proporţie din varianţa scorului observat şi ca varianţa erorii sistematice să fie minimă.

În practica aplicării testelor psihologice erorile pot apărea din următoarele surse:

- Din partea subiecţilor care participă la examinare, date fiind lipsa de motivare în obţinerea unui randament bun, concentrarea deficitară, starea precară a sănătăţii, oboseala accentuată. Erorile cauzate de factorii de natură fiziologică sunt printre cele mai imprevizibile şi mai individualizate. Deloc neglijabilă este sursa de eroare datorată lipsei de experienţă a subiecţilor în efectuarea unui test. Postura de participant la testare este relativ nouă pentru unii subiecţi, fapt ce poate duce la diminuarea randamentului. Trebuie să ţinem cont şi de existenţa „specialiştilor" care antrenează o serie de subiecţi pentru situaţia de testare psihologică. O specie aparte de erori din această categorie este dată de subiecţii care constituie un subgrup al grupului testat şi au o motivaţie diferită de cea a grupului participant la testare, situaţie ce poate să apară mai ales la testele de randament maximal.
- Din partea psihologilor care realizează testarea psihologică (printr-o serie de omisiuni în instrucţiunile de aplicare, prin limbajul utilizat, ambiguitatea enunţurilor) şi a operatorilor implicaţi în examinarea psihologică (explicaţii insuficiente, erori de codificare şi de corectare a răspunsurilor, erori în colectarea datelor). Variantele computerizate anulează, în majoritatea cazurilor, posibilitatea unor astfel de erori.

- Din partea testului însuşi. Din această categorie de erori fac parte cele datorate eşantionării itemilor care pot duce la versiuni lipsite de identitate ale unor teste considerate paralele, precum şi cele datorate impreciziilor de funcţionare şi de înregistrare, în cazul unor teste aparative, sau limitei neadecvate de timp.
- Din partea ambianţei în care se desfăşoară testarea psihologică, ea putând fi absolut improprie pentru o examinare.

Erorile datorate testului însuşi sunt mai uşor de prevenit, dar cele datorate subiecţilor sunt dificil de evitat. R.L. Thorndike s-a ocupat în mod special de studiul acestor erori. Existenţa mai multor surse de erori face ca autorii care au contribuţii de valoare în domeniul teoriei testului psihologic să formalizeze acest aspect, realizând o descompunere a varianţei erorii (Lord şi Novick, Gulliksen, Magnusson, Horst, Lienert, Bernier şi Pietrulewicz, Monica Albu şi Horia Pitaru)[1]. Astfel, H. Gulliksen descompune varianţa totală pentru un test oarecare în:

$$s_X^2 = s_V^2 + s_{EM}^2 + s_{VEQ}^2 + s_{EA}^2 + s_{EH}^2 + s_{ES}^2 + s_{VF}^2,$$

unde:

s_X^2 = varianţa totală;

s_V^2 = varianţa adevărată;

s_{EM}^2 = varianţa erorii de măsură;

s_{VEQ}^2 = varianţa adevărată specifică;

s_{EA}^2 = varianţa erorii datorate condiţiilor de administrare a testului;

s_{EH}^2 = varianţa erorii datorate hazardului în răspunsuri;

s_{ES}^2 = varianţa erorii datorate subiectivităţii în corecţie;

s_{VF}^2 = varianţa datorată fluctuaţiei în scorurile adevărate, de la o ocazie la alta (*apud* Bernier şi Pietrulewicz, 1997, p. 115).

David Magnusson caută să stabilească direcţiile din care ar putea fi afectată varianţa scorului total în cazul unei a doua aplicări a testului şi propune următoarea ecuaţie (1975, p. 114):

$$s_t^2 = s_T^2 + s_{e(m)}^2 + s_{e(adm)}^2 + s_{e(g)}^2 + s_{e(subj)}^2 + s_{T(fl)}^2,$$

unde:

s_t^2 = varianţa totală a testului;

s_T^2 = varianţa adevărată, care este ca atare;

$s_{e(m)}^2$ = varianţa erorii atribuite efectelor amintirii;

$s_{e(adm)}^2$ = varianţa erorii atribuite diversităţii modurilor de aplicare a testului;

1. Lucrarea noastră face referinţe insignifiante la demonstraţii matematice ale formulelor utilizate, având ca justificare argumentul că este inutil să redemonstrăm ceea ce a fost deja demonstrat. Cititorilor care doresc să parcurgă aceste demonstraţii, le recomandăm lucrările autorilor enumeraţi între paranteze. Vom utiliza, în cazul exemplificărilor, doar formulele finale ale unor astfel de demonstraţii.

$s^2_{e(g)}$ = varianţa erorii atribuite efectului explicaţiilor date pentru efectuarea probei;

$s^2_{e(subj)}$ = varianţa erorii atribuite obiectivităţii deficitare;

$s^2_{T(fl)}$ = varianţa adevărată, care exprimă fluctuaţia valorilor adevărate în timp.

5. Teoria probabilistă a testului psihologic

În primul rând, trebuie să specificăm ce semnificaţie are o teorie probabilistă şi care este înţelesul termenului „probabilist". Teoria măsurării precizează clar că, pentru a avea sens, un proces de măsurare trebuie să se bazeze pe un model şi trebuie să se raporteze la postulatele modelului. Într-o formă foarte generală de clasificare, modelele de măsurare pot fi deterministe sau probabiliste. Un alt criteriu de clasificare generală foarte răspândit se referă la numărul de dimensiuni (dincolo de nivelul nominal, rezultatele sunt dimensiuni), clasificarea rezultată distingând între modelele unidimensionale, care permit doar o dimensiune rezultantă, şi cele multidimensionale, care permit măsurări pe mai multe dimensiuni. Unidimensionalitatea poate fi tratată ca un caz particular al multidimensionalităţii. Teoria probabilistă a testului psihologic se bazează pe un model probabilist de măsurare. Facem o exemplificare cu modelul creat de Leon Guttman, numit *modelul ierarhic al lui Guttman* sau *modelul scalogramei unidimensionale.* Prezentăm un exemplu fictiv, cuprinzând rezultatele binare a 6 subiecţi (S_i) la 5 itemi (I_j), la o probă de raţionament deductiv (0 – răspuns incorect, 1 – răspuns corect).

Tabelul 4

Matrice de date observate şi matrice de date permutate

Matrice cu date observate

SUBIECŢI	ITEMI				
	I1	I2	I3	I4	I5
S1	0	0	1	0	1
S2	0	0	0	0	0
S3	0	0	1	1	1
S4	1	1	1	1	1
S5	0	1	1	1	1
S6	0	0	0	0	1

Matrice cu date permutate

SUBIECŢI	ITEMI				
	I5	I3	I4	I2	I1
S2	0	0	0	0	0
S6	1	0	0	0	0
S1	1	1	0	0	0
S3	1	1	1	0	0
S5	1	1	1	1	0
S4	1	1	1	1	1

La examinare, se presupune că nu toţi subiecţii au acelaşi potenţial în privinţa raţionamentului deductiv şi că nu toţi itemii au acelaşi grad de dificultate. Cercetătorii îşi propun în urma acestei examinări să ordoneze atât itemii (după dificultatea lor), cât şi subiecţii (după potenţialul lor). Cerinţa este de a ordona subiecţii în raport cu itemii şi invers, în aşa fel încât al doilea subiect plasat pe verticală să fie mai capabil decât primul, al treilea mai capabil decât al doilea ş.a.m.d., iar în acelaşi timp al doilea item pe orizontală să fie mai dificil decât primul, al treilea mai dificil decât al doilea ş.a.m.d.

Din matricea permutată se observă clar că un subiect care reuşeşte un anumit item trebuie să reuşească şi itemii de dificultate mai mică. Realizarea acestui obiectiv presupune enunţarea unor reguli de corespondenţă între rezultatele empirice şi rezultatul vizat. Matricea permutată poate fi transpusă geometric printr-un model al lui Van der Ven, care arată că, dacă o persoană a răspuns corect la un item, punctul reprezentând persoana pe o dreaptă ce traduce poziţia persoanelor şi a itemilor va fi poziţionat la dreapta punctului reprezentând respectivul item. În cazul în care punctul va fi situat la stânga unui item, aceasta semnifică faptul că nu a rezolvat corect itemul în cauză[1]. În figura următoare este prezentată diagrama Van der Ven pentru datele din matricea permutată anterioară.

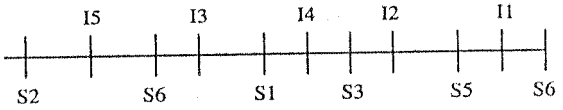


Figura 4. *Diagrama Van der Ven pentru datele din tabelul 4*

În exemplul nostru, fictiv, nu există nici o derogare de la model, datele conformându-se regulilor anunţate. Un coeficient de reproductibilitate ne poate arăta dacă un instrument de măsură construit după modelul Guttman satisface exigenţele sale teoretice. Se admite o toleranţă de 5% pentru devierea de la model pentru a pretinde că un instrument de măsură se revendică de la teoria lui Guttman. Rezultatul se consideră a fi înregistrat pe un acelaşi *continuum latent*. Exemplificarea a fost făcută pe unul dintre cele mai simple şi mai cunoscute modele (un model unidimensional, folosind date dihotomice). Modelele mai complexe cuprind mai multe matrice de date în care se folosesc diferite tipuri de date pentru diferite niveluri de măsură (nominal, ordinal etc.). Practic, modelele sunt construite pe baza unor ipoteze de structurare a datelor şi oferă proceduri de testare a îndeplinirii postulatelor lor.

Modelul lui Guttman este un model determinist şi implică următoarea supoziţie: dacă subiectul domină itemul („a domina" înseamnă a răspunde adecvat dificultăţii pe care o presupune, deci „a stăpâni" dificultatea respectivă), valoarea scalară a subiectului este superioară valorii scalare a itemului (ne referim la scala Van der Ven). În schimb, în optica probabilistă, se pleacă de la supoziţia că, dacă subiectul domină itemul, este *probabil* ca valoarea sa de scală să fie superioară aceleia a itemului.

Dacă se întocmesc curbe caracteristice ale itemilor (un graf al relaţiei dintre un *continuum latent* pe care itemul este destinat să-l măsoare şi probabilitatea de a furniza un răspuns corect), cea deterministă indică o trecere bruscă de la 0 la 1, în cazul în care competenţa creşte (funcţia este discontinuă), pe când cea probabilistă indică o relaţie monoton crescătoare. În cadrul opticii probabiliste, rezolvarea unui item mai greu şi ratarea unuia mai uşor nu constituie o derogare de la model. Trecerea de la optica deterministă la cea probabilistă nu înseamnă pur şi simplu tolerarea erorilor, ci gestionarea lor, adică înregistrarea şi estimarea influenţei acestora. Un model probabilist emite ipoteze asupra distribuţiei erorilor pentru a le putea estima, operaţiune omisă de modelele deterministe.

În opoziţie cu modelul clasic al testului psihologic, modelul probabilist se edifică pe ipoteza conform căreia trăsăturile psihice, în calitate de dimensiuni latente, nu pot fi situate

1. Adaptare după Paul Dickes, Jocelyne Tournois, André Flieler, Jean-Luc Kop, *La Psychométrie*, PUF, Paris, 1994, p. 140.

pe acelaşi plan cu comportamentul observabil. Autorii principali care s-au dedicat acestei teme sunt: G. Rasch (1960), G.H. Fischer (1968), P.F. Lazarsfeld şi N.W. Henry (1968), F.M. Lord şi M.R. Novick (1968), E.B. Andersen (1973), E.K. Scheuch şi H. Zehnpfenig (1974), J. Rost şi H. Spada (1978), R.K. Hambleton şi H.J. Rogers (1989), R.K. Hambleton, H. Swaminathan şi H.J. Rogers (1993), J. Grégoire (1994)[1]. Valoarea observată a testului îndeplineşte, în calitate de măsurare „deviată", funcţia unui indicator privind starea parametrului „latent trăit"; în consecinţă, valorile testului indică parametrul probabilist al caracteristicii. Dezvoltarea unui test cu ajutorul căruia trebuie să fie stabilit fundamentul pozitiv rezidual pe un continuum latent al caracteristicii presupune o teorie psihologică, raportată la o lege generală care guvernează raportul între comportamentul manifest în test şi o dimensiune latentă şi care permite determinarea intervalelor de încredere ale parametrului ce se cere estimat. Testul, în calitate de instrument de investigaţie, indiferent dacă este psihologic sau nu, duce implicit la concluzii cu caracter probabilist. Limbajul probabilist este o formă adecvată de exprimare în cazul studiilor vizând domenii de cercetare caracterizate de complexitate şi dinamism.

Deosebirile dintre optica probabilistă şi cea clasică, în privinţa măsurării, sunt ilustrate în figura următoare (adaptată după H. Benesch):

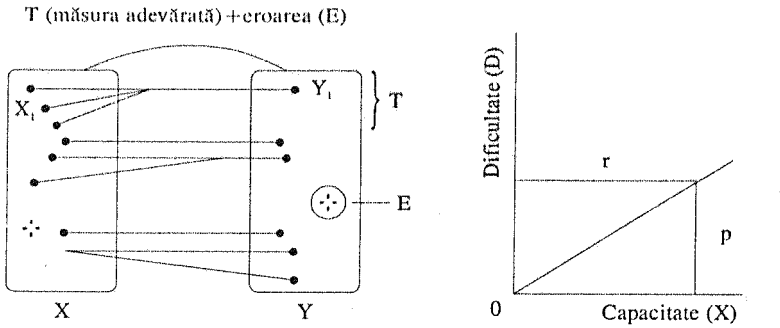


Figura 5. *Reprezentare spaţială pentru teoria clasică şi teoria probabilistă a testului psihologic*

Una dintre carenţele analizei clasice a itemului este dată de faptul că proprietăţile metrice ale acestuia sunt relative. Aproape toţi indicii pe care putem să-i calculăm depind de eşantionul de subiecţi utilizaţi, mai ales în domeniul cercetărilor de psihologie şcolară, unde problema privind relativitatea valorii itemilor se pune cu acuitate, deoarece examinarea cu acelaşi test de achiziţii şcolare (în momente diferite ale procesului de învăţământ) duce la oscilaţii serioase ale indicelui de dificultate. Cercetările noastre cu matricea progresivă Raven arată clare oscilaţii ale indicilor de dificultate a itemilor prin trecerea examinărilor de la mediul urban la mediul rural, de la studii generale la studii medii şi superioare. Crearea băncilor de itemi întâmpină, dată fiind oscilaţia indicelui de dificultate, serioase greutăţi şi se pune problema de a crea itemi independenţi de grupul de subiecţi pe care s-a calculat dificultatea. Mai ales în domeniul educaţiei se pune

1. Prezentăm între paranteze anii în care au apărut lucrări remarcabile pe această temă.

problema creării unor scale absolute independente de un anumit grup de referinţă. Desigur, aceasta pare o utopie, dar utopiile pot deveni realitate prin perseverenţa unor cercetători. Primele încercări în vederea creării unei scale care să nu depindă de un grup de referinţă au fost făcute de Guttman la începutul anilor '50, în scopul de a măsura atitudinile.

Modelul probabilist se sprijină pe postulatul potrivit căruia calitatea unui răspuns este determinată de trăsătura măsurată, dificultatea itemului şi discriminarea itemului. Deci probabilitatea de a reuşi la un item este o funcţie a caracteristicii subiectului (trăsătura măsurată) şi a proprietăţilor metrice ale itemului (dificultatea şi discriminarea sa). Psihometricienii au iniţiat diverse modele de relaţii funcţionale între item şi caracteristica măsurată. Un element comun acestor modele diverse este faptul că ele împărtăşesc postulatul după care totalitatea itemilor unui test măsoară aceeaşi caracteristică psihologică, dar nivelul răspunsurilor la acest item poate fi afectat de erori aleatorii. Obiectivul unor astfel de modele este să permită, pe de o parte, o estimare a proprietăţilor metrice ale itemilor invarianţi pe suprafaţa unei populaţii şi, pe de altă parte, o estimare a unor trăsături psihologice independente de itemii utilizaţi în măsurare. Prin denumirea de MRI (modele de răspuns la itemi) sunt reprezentate majoritatea modelelor probabiliste.

Postulatul fundamental al MRI susţine că performanţa unui subiect la un item poate fi explicată printr-un factor numit *trăsătură latentă*, care poate fi o trăsătură de personalitate, o aptitudine cognitivă, o capacitate şcolară. Relaţiile dintre performanţa la item şi trăsătura latentă poate fi prezentată prin intermediul unei funcţii numite *funcţia caracteristică a itemului* sau, folosind noţiunea cea mai frecventă, *curba caracteristică a itemului* (CCI). Prezentăm în figura următoare situaţia a trei itemi având diferite dificultăţi şi puteri discriminatorii:

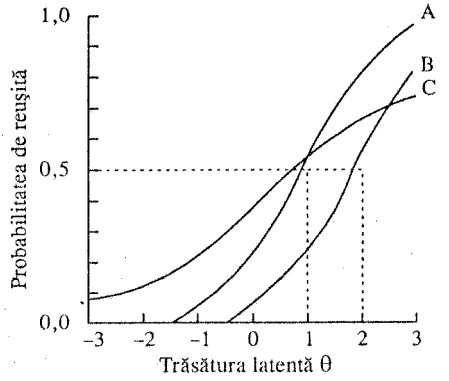


Figura 6. *Curba caracteristică a trei itemi de dificultăţi diferite*

Această curbă ia forma unui S puţin mai alungit, în care trăsătura latentă apare în abscisă şi este reprezentată prin convenţie cu litera grecească *theta* (θ). Trăsătura latentă va fi distribuită normal în interiorul populaţiei, gradaţiile abscisei corespunzând valorilor standardizate z.

Pe acelaşi grafic, posibilitatea de a da un răspuns corect apare în ordonată. Valoarea în acest caz oscilează între 0 şi 1. Pe măsură ce un subiect se situează la un nivel mai ridicat pe trăsătura latentă, probabilitatea de a răspunde corect la item creşte. Dar probabilitatea

de reuşită depinde la fel de mult de dificultatea itemului. La valori egale ale lui θ, valorile probabilităţii de reuşită la item cresc sau scad după acest nivel de dificultate.

Prin convenţie, valoarea reprezentând dificultatea unui item este egală cu valoarea lui θ pentru care probabilitatea de a da un răspuns corect este egală cu 0,50. Nivelul de dificultate al unui item reprezintă primul parametru permiţând descrierea CCI a acestui item.

Însă indicele de dificultate nu este singurul parametru în joc, existând şi un parametru secund, care este puterea discriminativă a itemului. Discriminarea unui item este reprezentată de panta CCI (înclinarea curbei). Cu cât panta este mai abruptă, cu atât itemul este mai discriminativ.

Primele lucrări ştiinţifice de specialitate referitoare la funcţia care leagă un item de trăsătura latentă au apărut la începutul anilor '50, remarcabile fiind cele ale lui F.M. Lord. Primele aplicaţii au fost efectuate pe testele de inteligenţă. Dar evoluţiile teoretice şi aplicaţiile MRI au fost impulsionate în mod deosebit prin publicarea, în 1960, a unui studiu al matematicianului danez Georg Rasch, intitulat *Probabilistic models for some intelligence and attainment tests*. Rasch a fost primul cercetător care a utilizat o funcţie logistică pentru a analiza datele ce apar în cadrul construcţiei testului psihologic. Totuşi, răspândirea modelului în rândul psihologilor datorează foarte mult lucrărilor de specialitate ale lui G.H. Fischer, care îl matematizează după unele exigenţe înalte. Modelul propus de Rasch este cel mai simplu din MRI şi se sprijină pe postulatul că toţi itemii discriminează în mod egal şi, în consecinţă, singurul parametru care urmează să fie estimat este dificultatea itemului. Direcţia deschisă de matematicianul danez în construcţia testului cunoaşte un dinamism deosebit, fiind completată de contribuţii consistente[1]. Modelul parametrului utilizat de Rasch, numit adesea şi *modelul lui Rasch*, susţine că probabilitatea de a reuşi la un item poate fi estimată prin formula următoare[2]:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}},$$

unde:

$P_i(\theta)$ = probabilitatea ca un subiect oarecare, posedând o aptitudine, să răspundă corect la itemul i;
e = constanta lui Neper, valoare care corespunde numărului iraţional 2,718281;
b_i = parametrul de dificultate al itemului i.

Modelul lui Rasch are o restricţie importantă, şi anume aceea că toţi itemii unui test posedă aceeaşi capacitate de discriminare. Exigenţa respectivă poate fi întâlnită atunci când itemii sunt foarte asemănători, în testele de achiziţii şcolare focalizate pe un domeniu precis. Dar, în multe alte cazuri, cerinţa nu este uşor de satisfăcut. Din aceste raţiuni, un model care ia în considerare diversitatea puterii discriminative a itemilor este din ce în ce mai utilizat. Modelul logistic cu doi parametri (dificultate şi discriminare) a fost dezvoltat de A. Birnbaum, în 1968.

1. Un studiu sintetic relativ recent al direcţiilor de dezvoltare a modelului Rasch este realizat de Jürgen Rost, „Was ist aus dem Rasch-Model geworden?", în *Psychologische Rundschau*, Hogrefe, 50, 3/1999, pp. 140-156.
2. G. Rasch şi G.H. Fischer au introdus literele greceşti în exprimarea formulelor. W. Guthjahr creează variante de formule cu litere latine. Explicaţii suplimentare pot fi găsite în lucrarea lui Guthke, Böttcher şi Sprung, *Psychodiagnostik*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1990.

Conform acestui model, probabilitatea de reuşită la un item poate fi estimată printr-o extensie a formulei lui Rasch:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}},$$

unde:
a_i = parametrul de discriminare al itemului i, proporţional pantei curbei b_i;
D = factor de scalare (*scaling factor*) egal cu 1,7.

Valoarea lui a_i se situează în mod obişnuit între 0 şi 2, dar când această valoare este negativă înseamnă că probabilitatea de a reuşi itemul se diminuează în funcţie de capacitatea subiectului. O astfel de situaţie nu are deloc sens din punctul de vedere al practicii testului psihologic clasic, deoarece un item care prezintă un parametru de discriminare negativ este în mod obişnuit eliminat sau, cel puţin, revizuit. De altfel, situaţia de a întâlni itemi a căror valoare a a_i este superioară lui 2 este foarte rară (o astfel de valoare indică o pantă foarte abruptă). Realitatea unor inevitabile erori de măsură face puţin probabilă observarea unei discriminări mai marcante.

Majoritatea cercetărilor asupra MRI au fost realizate cu itemi având dispozitive de răspuns de tipul „adevărat - fals". Când se realizează o astfel de modalitate, există riscul ca subiectul, care nu posedă nici o abilitate, să reuşească un item prin hazard. În acest caz, asimptota cea mai joasă a CCI este net superioară lui 0. Pentru a face faţă unei astfel de eventualităţi, s-a propus includerea unui al treilea parametru în ecuaţia ultimă: parametrul „pseudo-şansei" (o denumire care poate să mire). R.K. Hambleton, H. Swaminathan şi H.J. Rogers susţin că nu este corect să fie numit „parametrul şansei", deoarece valorile acestui parametru sunt inferioare celor cărora le corespunde o alegere total aleatorie.

Următoarea ecuaţie corespunde modelului cu trei parametri:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}},$$

unde c_i este parametrul „pseudo-şansei".

Curba caracteristică a unui item după modelul logistic cu trei parametri este prezentată în figura următoare:

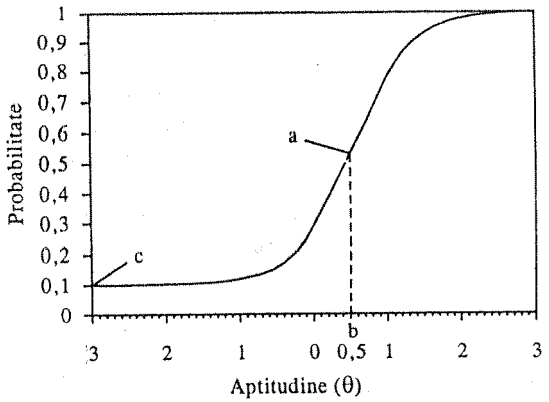


Figura 7. *Curba caracteristică a unui item (conform modelului logistic cu trei parametri)*

Avantajul major al MRI este că ele permit determinarea parametrilor caracteristici ai unui item (dificultate, discriminare, pseudo-şansă) independent de caracteristicile subiecţilor care au permis estimarea. Aceste proprietăţi de invarianţă a parametrilor fac posibilă folosirea unei bănci de itemi. Banca de itemi este un vast ansamblu de itemi din care se poate construi un test. Nu este o realizare strictă a opticii probabiliste de abordare a construcţiei testului. Crearea unui ansamblu de itemi presupune un studiu empiric în care se schimbă subiecţii şi momentul de aplicare. Itemii compunând băncile de itemi posedă caracteristici metrice care nu sunt comparabile. Idealul în construcţie este să se obţină caracteristici ale itemilor independente de grupul de subiecţi care au permis să fie calculate. Construirea scalelor de măsură absolut independente de grupul de referinţă (scale absolute) presupune că un scor total nu este dependent de grupul de referinţă. Practicienii care doresc să alcătuiască un test pot selecţiona itemii care nu au fost analizaţi cu acelaşi eşantion de subiecţi, fapt ce oferă posibilitatea unei mulţimi de combinaţii de itemi. În mod particular, este interesantă situaţia în care cineva vrea să retesteze subiecţii evitând efectul repetiţiei aceloraşi itemi. Un alt avantaj constă în posibilitatea de a compara subiecţii care au trecut ansambluri de itemi diferiţi (ca în cazul *testing*-ului adaptativ).

Invarianţa parametrilor poate să-i mire pe practicienii obişnuiţi cu relativitatea analizei clasice. O comparaţie poate să ajute la înţelegerea acestei proprietăţi. În modelul regresiei liniare, relaţia dintre o variabilă X şi o variabilă Y este reprezentată printr-o dreaptă de regresie. Această dreaptă de regresie este descrisă printr-o ecuaţie în care parametrii sunt panta şi ordonata la origine, parametri estimaţi plecând de la un ansamblu restrâns de valori ale lui X. Dacă modelul de regresie liniară este adecvat, orice alt ansamblu de valori X şi Y trebuie să estimeze aceleaşi valori ale parametrilor. Este logic să considerăm că această proprietate de invarianţă a parametrilor se manifestă şi la MRI, care pot fi considerate modele de regresie nonliniară. Cu toate că MRI par foarte seducătoare la prima abordare, practicianul nu trebuie să piardă din vedere faptul că ele se bazează pe postulate foarte puternice, adică greu de satisfăcut. Înainte de a utiliza MRI, este necesar să se verifice dacă anumite exigenţe sunt satisfăcute la nivelul itemilor şi al realităţilor pe care aceştia le măsoară. MRI se sprijină pe două postulate: unidimensionalitatea şi independenţa locală.

Exigenţa de unidimensionalitate presupune că toţi itemii unui test trebuie să măsoare una şi aceeaşi trăsătură. În practică, acest criteriu nu este aproape niciodată îndeplinit, date fiind inevitabilele erori de măsură şi complexitatea trăsăturilor măsurate. Când exigenţa unidimensionalităţii nu poate fi satisfăcută, MRI nu pot fi utilizate, situaţie care face ca actualmente să fie în curs de elaborare MRI multidimensionale. Realizarea lor este destul de complicată, comportând numeroase constrângeri metodologice, printre care şi mărimea eşantionului.

Exigenţa de independenţă locală presupune că trăsătura ce face obiectul evaluării trebuie să fie singurul factor care determină variabilitatea răspunsurilor la itemii testului. O dată ce trăsătura măsurată a fost luată în considerare, nu trebuie să existe nici o relaţie între răspunsurile unui subiect la diferiţi itemi. Atunci când consemnul unui test dă informaţii, ce permit să se răspundă mai uşor la anumiţi itemi, exigenţele de independenţă locală nu sunt respectate. Unii subiecţi pot să remarce aceste informaţii, alţii nu. Drept consecinţă, scorul testului nu conţine anumiţi indici utili. Exigenţa de independenţă locală nu mai este satisfăcută dacă, într-un test de matematică, anumiţi itemi fac apel,

din întâmplare, şi la cunoştinţe de geografie. Reuşita la acest item nu este determinată de singura trăsătură latentă pe care noi dorim să o măsurăm. În comparaţie cu ceilalţi, subiecţii cu bune cunoştinţe în geografie vor prezenta o mai mare probabilitate să obţină un scor ridicat la respectivul test de matematică.

Când exigenţa de unidimensionalitate este satisfăcută, exigenţa de independenţă locală este şi ea satisfăcută. Relaţia inversă nu este posibilă. Se pot observa un spaţiu latent multidimensional şi, în acelaşi timp, o independenţă locală a itemilor testului. Această situaţie se produce când un al doilea factor influenţează toţi itemii în mod egal. De exemplu, într-un test de fizică, subiecţii pot să aibă de citit un scurt enunţ şi, dacă toţi subiecţii au acelaşi potenţial de lecturare (în sensul de comprehensiune verbal-logică a problemei), nu se va produce diferenţierea. Atunci va exista independenţă locală şi subiecţii se vor distinge după competenţa lor în fizică. Testul nu va putea fi apreciat ca unidimensional pentru că performanţele vor fi susţinute prin cel puţin doi factori, al doilea reprezentându-l potenţialul de lectură.

MRI implică anumite constrângeri metodologice. În mod clar, utilizarea practică a acestor modele este mai complexă decât acele tehnici rezultate din teoria clasică a testului. Practicienilor li se cer serioase competenţe teoretice şi aparatură informatică puternică, limitându-se astfel câmpul de aplicare al MRI. Printre constrângerile metodologice, cele mai importante sunt *estimarea parametrilor* şi *procedurile de legătură*. Prima constrângere metodologică, estimarea parametrilor, mai este numită şi „calibrarea" itemilor. Demersul este relativ complex şi apelul la calculator este aproape întotdeauna necesar, mai ales dacă numărul de itemi şi de subiecţi necesari pentru o calibrare precisă a parametrilor este destul de ridicat. Pe piaţă există un număr apreciabil de programe de calibrare a itemilor pentru MRI, cu doi sau trei parametri.

Crearea băncilor de itemi a venit, în parte, ca o consecinţă a constrângerilor în estimarea parametrilor. Cercetătorii cu preocupări în acest domeniu de vârf (în speţă Van der Linden) susţin argumentat că o bancă de itemi fără MRI este irealizabilă. Determinările parametrilor nu se fac direct, ci prin aproximări succesive. La fiecare utilizare a itemului, rezultatele sunt încorporate în ansamblul rezultatelor anterioare, permiţând ameliorarea calibrării itemului în chestiune. Există astfel o relaţie reciprocă şi dinamică între băncile de itemi şi testele construite, care se bazează informaţional pe acestea.

A doua constângere metodologică o reprezintă procedura de legătură, desemnată prin termenul englezesc *linking* în literatura de specialitate. În cadrul MRI, parametrii itemilor sunt invarianţi, dar în practică acest lucru nu se realizează întotdeauna. În afara procedurii de estimare a parametrilor, punctul 0 de estimare a dificultăţii itemilor este arbitrar centrat pe media estimărilor lui θ (θ desemnează capacitatea subiectului), pentru grupul de subiecţi implicaţi în analiză. Deoarece poziţia 0 variază după grup, parametrii obţinuţi sunt *ipso facto* relativi. Această problemă poate fi rezolvată destul de uşor, deoarece parametrii unui item sunt efectiv invarianţi ţinând cont de o transformare liniară care permite ca toţi să fie plasaţi pe aceeaşi scală. În alţi termeni, în afara calibrării itemilor, este necesară determinarea unei constante care permite ca valorile obţinute să fie transformate şi ajustate la o scală de referinţă. Pentru a realiza această procedură de legătură, sunt posibile mai multe tehnici:

- plasarea itemilor comuni în diferite ansambluri de itemi;
- utilizarea subiecţilor comuni în timpul efectuării diverselor grupe de itemi;
- efectuarea simultană a celor două operaţiuni.

Estimarea parametrilor. Este o operaţie crucială şi, în funcţie de calitatea acestei estimări, se conferă sens utilizării MRI. Interesul pentru MRI va fi pierdut dacă nu se obţine invarianţa parametrilor. După primele lucrări ale lui Rasch, au fost propuse numeroase proceduri de estimare a parametrilor, cea mai utilizată astăzi fiind metoda verosimilităţii maxime (*marginal maximum likelihood*), în speţă în programele informatice de dată recentă : BILOG (autori : Mislevy şi Bock) sau XCALIBRE (firma producătoare : Assessment Systems Corporation, 1994).

Când vrem să estimăm parametrii unui ansamblu de itemi, răspunsurile subiecţilor sunt singurele informaţii de care dispunem. Suntem atunci constrânşi să estimăm în acelaşi timp trăsătura θ a subiecţilor şi parametrii itemilor. Realizarea acestei duble estimări este departe de a fi evidentă. Pentru a înţelege logica unei asemenea situaţii, este necesar să se plece de la o situaţie mai simplă în care parametrii sunt cunoscuţi şi în care trebuie să fie estimată o singură trăsătură θ a subiecţilor pe baza modelului rezultatelor. Probabilitatea ca un subiect, posedând o capacitate θ, să obţină un răspuns U_j ($U_j = 1$, dacă răspunsul este corect, şi 0, dacă este fals) se notează $P(U_j / \theta)$. În virtutea postulatului de independenţă locală, probabilitatea de a observa un model de răspuns la un ansamblu de n itemi este egală cu produsul probabilităţilor de răspuns la fiecare dintre itemii respectivi :

$$P(U_1, U_2, ..., U_n / \theta) = \Pi P(U_j / \theta)$$

Formula se poate scrie :

$$P(U_1, U_2, ..., U_n / \theta) = \Pi P_j^{U_j} Q_j^{1-U_j}$$

$$P_j = P(U_j / \theta)$$

$$Q_j = 1 - P(U_j / \theta)$$

Dacă în această formulă se plasează modelul de răspuns efectiv observat, ea nu poate fi interpretată în manieră probabilistă. Funcţia pe care o obţinem astfel este numită *funcţie de verosimilitate* :

$$L(u_1, u_2, ..., u_n / \theta) = \Pi P_j^{U_j} Q_j^{1-U_j}$$

Estimarea trăsăturii θ a unui subiect constă în posibilitatea de a calcula valoarea θ care maximizează funcţia de verosimilitate. Pentru a găsi valoarea maximă a acestei funcţii se foloseşte o procedură de aproximare succesivă, cea mai frecvent utilizată fiind aceea a lui Newton-Raphson. Metoda nu se poate aplica atunci când răspunsurile unui subiect sunt în întregime corecte sau în întregime eronate. În acest caz, estimarea care maximizează funcţia de verosimilitate este :

$$\theta = +\infty \text{ sau } \theta = -\infty.$$

Când nu cunoaştem nici valorile lui θ, nici valorile parametrilor, situaţia este şi mai complexă, deoarece trebuie să considerăm în acelaşi timp ansamblul a n itemi ai testului şi modelele de răspuns a N subiecţi care au răspuns la respectivii itemi. În acest caz, funcţia de verosimilitate se scrie :

$$L(u_1, u_2, ..., u_j, ..., u_n / \theta, a, b, c) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}}$$

În formulă, s-a considerat cazul în care trebuie să fie estimaţi trei parametri (a, b şi c). Aceeaşi formulă poate fi adaptată pentru cazul în care numai unul sau doi parametri trebuie să fie estimaţi.

Pentru a determina valoarea lui θ şi parametrii itemilor care maximizează funcţia de verosimilitate pot fi urmate două metode : estimarea verosimilităţii maxime legate (sau condiţionate) şi estimarea verosimilităţii maxime marginale (sau necondiţionate).

Metoda verosimilităţii maxime legate prezintă anumite carenţe. Prima ar consta în aceea că parametrii itemilor eşuaţi sau reuşiţi de către toţi subiecţii nu pot fi estimaţi. La fel merg lucrurile pentru valorile lui θ atunci când subiecţii obţin numai răspunsuri corecte sau numai răspunsuri false. Din acest motiv, programele care utilizează procedura verosimilităţii maxime legate elimină dintr-o dată subiecţii şi itemii ale căror scoruri sunt uniform 1 sau uniform 0. O a doua carenţă a metodei apare la modelele cu doi sau trei parametri, pentru care estimările sunt instabile, dacă nu se utilizează un număr mare de subiecţi sau itemi. Din acest ultim motiv, metoda verosimilităţii maxime legate nu mai este utilizată astăzi decât în programele care realizează analiza după modelul lui Rasch (de exemplu, RASCAL, *Assessment Systems Corporation*). În alte cazuri, se preferă *metoda verosimilităţii maxime marginale*. Această metodă este mult mai costisitoare în calcule decât precedenta, dar permite obţinerea unor estimări mai stabile pentru modelele cu doi sau trei parametri.

Oricare ar fi procedura utilizată, caracteristicile eşantionului de subiecţi joacă un rol important în calitatea estimării parametrului itemilor. În particular, un eşantion omogen de subiecţi antrenează estimări instabile ale parametrilor modelului (Hambleton, 1994). În scopul de a garanta eterogenitatea şi de a reduce impactul erorilor de măsură, este necesar un eşantion de subiecţi destul de mare. C.L. Hulin şi colaboratorii săi au evaluat mărimea optimală a acestui eşantion pentru modelele cu doi sau trei parametri. Astfel, ei au generat date simulate pe eşantioane de 200, 500, 1.000 şi 2.000 de subiecţi, la teste de 15, 30 şi 60 de itemi. Datele au fost analizate cu programul LOGIST. Rezultă, pentru modelul cu doi parametri, că un test de 30 de itemi şi un eşantion de 500 de subiecţi permit obţinerea unor estimări ale parametrilor relativ stabile. Cu modelul de trei parametri, acelaşi obiectiv poate fi atins cu un test de 60 de itemi şi un eşantion de 1.000 de subiecţi. În cazul modelului cu un parametru, cercetătorii americani B.D. Wright şi M.H. Stone (1979) recomandă utilizarea a cel puţin 20 de itemi şi a unui eşantion de 200 de subiecţi pentru a obţine o estimare satisfăcătoare. Estimarea parametrilor este un procedeu relativ costisitor din punctul de vedere al timpului necesar, deoarece presupune un proces major de colectare a datelor. Exigenţa respectivă constituie o limită reală pentru aplicarea MRI de o complexitate crescută.

În practică, datele utilizate pentru estimarea parametrilor nu se ajustează niciodată perfect la modelul ales. Diverse teste statistice au fost puse la punct pentru a evalua gradul de ajustare a modelului. Când această adecvare este slabă, nu putem obţine estimări invariante ale parametrilor. Itemii inadecvat ajustaţi trebuie să fie, în consecinţă, îndepărtaţi. Totodată, se cere prudenţă înainte de a respinge un item, deoarece testele de ajustare sunt foarte sensibile la mărimea eşantionului. Când eşantionul este mic, unele probleme de ajustare relativ importante pot să nu fie detectate, dar, atunci când eşantionul este foarte mare, problemele de ajustare minore riscă să conducă la respingerea itemului incriminat. În acest sens, ca un exemplu, prezentăm în continuare rezultatele lui Hambleton şi Murray (1983, *apud* Laveault şi Grégoire, 1997, p. 299). Datele au fost analizate cu

ajutorul programului BICAL (Wright şi Stone, 1979). Acest program permite să se realizeze o analiză de itemi după modelul lui Rasch cu ajutorul procedurii verosimilităţii maxime legate. Programul calculează, de asemenea, valorile lui *t* pentru a detecta itemii inadecvat ajustaţi la model, cu prag între 0,01 şi 0,05.

Mărimea eşantionului	Numărul de itemi inadecvat ajustaţi	
	$p < 0,05$	$p < 0,01$
150	20	5
300	25	17
600	30	18
1.200	38	28
6.000	42	38

Problema pe care vrem să o subliniem nu este specifică programului informatic, nici unui test de ajustare. Este vorba despre o problemă generală. În 1985, Hambleton şi Swaminathan sugerează că nu trebuie să ne îndreptăm atenţia asupra testelor de ajustare şi recomandă o investigaţie mai largă, privind adecvarea dintre model şi date. Concluziile trebuie să rezulte, astfel, din convergenţa unui ansamblu de indici. Trei categorii de indici trebuie să reţină atenţia practicianului (după R.K. Hambleton şi colaboratorii săi) :

1. Informaţiile cuprinzând validitatea postulatelor modelului utilizat pentru analiza datelor (rezultatele analizelor privind unidimensionalitatea ansamblului itemilor fac parte dintre acestea).
2. Informaţiile cuprinzând proprietăţile aşteptate pe baza modelului. Astfel se poate evalua dacă este confirmată proprietatea de invarianţă a parametrilor, comparând parametrii obţinuţi pe mai multe eşantioane de subiecţi ai populaţiei.
3. Informaţiile privind predicţiile realizate pe baza modelului (de exemplu, se poate stabili diferenţa dintre performanţele efective ale unui grup de subiecţi la un item şi acelea care au putut fi prezise pe baza nivelului de aptitudini – valoarea lui θ – al aceloraşi subiecţi).

Trebuie precizat că modelul probabilist nu este actualmente nici pe departe la fel de dezvoltat ca modelul clasic al testului psihologic. Probele scalate după modelul ştiinţific impus de Rasch sunt totuşi nişte rarităţi. Rasch şi-a pus în practică modelul încă din 1960, prin două baterii de teste de inteligenţă, formate numai din două subteste, care au fost utilizate de către armata daneză (teste apropiate de matricea progresivă Raven)[1]. J. Wendeler construieşte în 1968 un „test de calcul", iar cercetătorii germani W. Conrad, U. Baumann şi V. Mohr creează *Mannheimer Test zur Erfassung des physikalisch-technischen Problemlosen*. Tot cu ajutorul modelului probabilistic al testului, K.O. Kubinger şi E. Wurst examinează omogenitatea testului AID (*Adaptativen Intelligenzdiagnostikum*). Exemplele nu au epuizat realizările în acest sens.

1. Informaţii pe această temă pot fi găsite în lucrarea lui Guthke, Böttcher şi Sprung, *Psychodiagnostik*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1990, p. 162.

6. Criteriile de apreciere a unui test

6.1. Consideraţii preliminare. Controverse

Înainte ca datele obţinute prin aplicarea unui test psihologic să poată fi folosite în activităţi diferenţial psihodiagnostice, este necesar să se stabilească dacă acestea îndeplinesc anumite condiţii. În primul rând, trebuie să fim siguri că măsurăm cu precizie (valoarea indicată de instrumentul de măsură este reproductibilă la a doua aplicare a testului la aceeaşi persoană) şi că măsurăm trăsătura sau caracteristica precizată anterior folosirii testului, şi nu alta. Condiţiile de valabilitate a unui test trebuie îndeplinite în mare parte empiric, adică prin aplicări şi reaplicări impuse de nevoi psihodiagnostice concrete.

În privinţa stabilirii numărului de condiţii sau criterii, menţionăm că nu există o unanimitate de opinie, deosebiri apărând atât cu privire la denumirile criteriilor, cât şi la ordonarea şi ierarhizarea lor. Există optici restrânse şi optici extinse de abordare a problematicii criteriilor de apreciere a testului psihologic. Prezentăm în continuare punctele de vedere ale unor autori de specialitate, cu preocupări în acest domeniu.

Serge Chamboulant, specialist în psihologia aplicată, precizează calităţile necesare ale testului, ca şi ale oricărei măsurări, indiferent dacă acestea sunt fizice sau biometrice. Conform exigenţelor exprimate de S. Chamboulant (1951, p. 26), un test trebuie să prezinte noutate pentru subiect (exigenţă nu uşor de satisfăcut, mai ales în ţările în care experienţa testării psihologice este curentă), să fie interesant şi rapid, obiectiv în aplicaţie şi corectare (deci independent de calităţile examinatorului), gradabil şi sensibil pentru a putea măsura cele mai infime diferenţe între indivizi, fidel (adică să dea întotdeauna aceleaşi note sau unele foarte apropiate) şi valid (adică să-şi îndeplinească rolul de prognostic pentru succesul într-o anumită activitate).

Autorul englez David Bartram (1994) acordă, într-un articol dintr-o cunoscută lucrare destinată construcţiei testului psihologic, o atenţie aproape exclusivă fidelităţii şi validităţii, abia amintind de existenţa altor criterii.

Autorul german Jürgen Guthke, amintit anterior, consideră că printre calităţile testului psihologic figurează concordanţa interpersonală sau obiectivitatea, validitatea sau valabilitatea, fidelitatea sau precizia de măsurare, sensibilitatea şi, în final, normativitatea (vezi Guthke, Böttcher şi Sprung, 1990).

În bine cunoscutul său tratat *Testaufbau und Testtheorie* (1967) consacrat analizei şi construcţiei testului psihologic, G. Lienert întocmeşte (p. 12) o clasificare în care grupează criteriile de apreciere (*Gutenkriterien*) în principale şi secundare. Dintre criteriile principale fac parte obiectivitatea, fidelitatea (*Reliabilität*) şi validitatea, iar dintre cele secundare normativitatea, comparativitatea, economicitatea şi utilitatea. Considerăm că această clasificare este pertinentă pentru o cercetare sistematică a testului psihologic.

Ursula Şchiopu (1997, p. 692) apreciază un număr redus de calităţi, şi anume validitatea, omogenitatea sau fidelitatea şi sensibilitatea.

Peter R. Hofstätter afirmă că în faţa testelor se ridică trei exigenţe : a) diferenţierea subiecţilor unii de alţii ; b) realizarea acestei diferenţieri în mod precis şi c) realizarea ei într-un mod relevant din punct de vedere diagnostic. Prima cerinţă desemnează normativitatea sau standardizarea drept criterii de apreciere, a doua fidelitatea, iar a treia validitatea sau valenţa diagnostică. Punerea semnului de egalitate între normativitate şi standardizare

este, cu siguranţă, o forţare în interpretarea noţiunilor (în acest sens, vezi Hofstätter, 1972, p. 292). Aspectele la care se referă standardizarea sunt mai diverse decât cele în care se operează cu norme.

Claude Duhamel (1992, pp. 34-37) este foarte succint în prezentarea calităţilor testului, enumerând etalonarea, sensibilitatea şi validitatea, lucrarea sa fiind printre cele relativ moderne de prezentare a testelor care se folosesc în recrutarea personalului.

Wilhelm Hehlmann (1968, pp. 580-581) consideră că valoarea diagnostică a unui test depinde de standardizarea, fidelitatea şi validitatea sa.

În una dintre lucrările de referinţă din literatura de specialitate românească asupra metodelor de psihodiagnostic, Mariana Roşca (1972, pp. 22-28) descrie calităţile testelor, şi anume standardizarea, etalonarea (care stau la baza construcţiei testelor), fidelitatea, validitatea şi sensibilitatea.

Într-o lucrare recentă referitoare la testele de tip *screening*, Nicolae Mitrofan (1997, pp. 20-21) prezintă o serie de caracteristici ale acestora, şi anume etalonarea, economicitatea, fidelitatea, validitatea, utilitatea şi standardizarea. În cadrul prezentării testelor pentru examinarea copilului mic, se face referire mai ales la validitate şi fidelitate.

Heiner Legewie şi Wolfram Ehlers (1978, pp. 20-21) consideră ca fiind *Gutenkriterien* obiectivitatea, fidelitatea şi validitatea.

O foarte interesantă prezentare a calităţilor testului psihologic realizează Lothar şi Helga Sprung (1987, pp. 147-210), studiul lor fiind consacrat metodicii şi metodologiei psihologice. Criteriile de apreciere, care sunt atribuite testului, fac parte şi din sistemul metodicii psihologice în general, fiind enumerate validitatea, fidelitatea, concordanţa interpersonală sau obiectivitatea, normativitatea şi utilitatea. Fiecare dintre aceste calităţi suportă influenţa unei pedanterii taxonomice rar întâlnite la alţi autori.

Monica Albu şi Horia Pitariu (1993, pp. 102-103) introduc printre calităţile testelor (cu referire specială la testele de cunoştinţe), pe lângă validitate, fidelitate, obiectivitate, putere discriminativă, aplicabilitate, şi o serie de denumiri rar întâlnite la alţi autori, şi anume:

- caracterul reprezentativ al testului (cu referire la itemii aleşi pentru a compune un test);
- pertinenţa testului (gradul de respectare a criteriilor stabilite pentru selectarea întrebărilor în aşa fel încât acestea să se conformeze scopului instrumentului de măsură);
- specificitatea testului (calitate dobândită în situaţia în care un subiect, nedispunând de cunoştinţele ce fac obiectul verificării, obţine un scor egal cu cel pe care l-ar fi realizat dacă ar fi răspuns la întâmplare).

Un termen similar pentru „fidelitate" este cel de „fiabilitate", concept folosit exclusiv în argumentaţie de Dany Laveault şi Jacques Grégoire (1997) şi, ocazional, de J.J. Bernier şi B. Pietrulewicz (1997). Sensul este, credem, o transpunere a celui din tehnică, adică probabilitatea bunei funcţionări a unei componente sau a unui aparat într-o perioadă de timp dată. Urmărind desfăşurarea de proceduri demonstrative şi argumentarea teoretică a autorilor, considerăm că suntem în prezenţa unei extensii semantice bine venite a ceea ce se denumise până acum fidelitate.

Vom căuta să abordăm fiecare noţiune considerată de autorii de specialitate ca fiind criteriu de apreciere sau de calitate a testului (însumând acel ansamblu de trăsături care în subsidiar se sprijină pe un set clar precizat de proceduri statistico-matematice),

satisfacerea sau îndeplinirea ansamblului format de acestea conferind identitate metodei testului. Variaţiile numărului de calităţi ale testului în abordările diverşilor autori au două cauze: 1) extinderea semnificaţiei denumirii de calitate sau criteriu de apreciere a testului asupra unor elemente contextuale şi conjuncturale care influenţează accidental, sporadic, valorile rezultate dintr-o examinare şi 2) preocuparea superficială pentru criteriile de apreciere în lucrări care aveau în atenţie alte aspecte ale testului.

Dintre autorii de marcă preocupaţi de precizarea conţinutului şi fundamentarea criteriilor de apreciere a testelor psihologice, îi enumerăm pe H. Gulliksen, F.M. Lord şi M.R. Novick, D. Magnusson, G. Lienert, P. Horst, D. Laveault şi J. Grégoire, J. Guthke, L. Sprung, L. Michel, M. Reuchlin, J. Tournois, P. Dickes, L.J. Cronbach, G.C. Gleser, A. Anastasi, R.B. Cattell, C. Burt, P. Kline şi J.M. Faverge. Aceştia se ocupă preponderent de două criterii de apreciere, considerate a fi cele mai importante: fidelitatea şi validitatea. Dacă aceste două calităţi nu sunt asigurate, nu are nici un rost să ne ocupăm de modul de satisfacere a celorlalte (pentru care există anumite tehnici constructive de ameliorare). Nici un fel de clasificare a criteriilor de valabilitate a testului psihologic nu omite fidelitatea şi vaditatea. Considerăm că obiectivitatea este omisă pe nedrept dintre cele mai importante criterii. Ca atare, vom aborda pentru început această calitate a testului psihologic.

6.2. Criterii principale de apreciere a testului psihologic

6.2.1. *Obiectivitatea sau concordanţa interpersonală a testului psihologic*

Obiectivitatea sau concordanţa interpersonală reprezintă gradul în care rezultatele unui test psihologic sunt independente de persoana examinatorului care exploatează psihodiagnostic acest instrument. Putem spune că un test este deplin obiectiv atunci când diverşi examinatori obţin aceleaşi rezultate la acelaşi subiect. Desigur, în aplicările practice, se obţine foarte rar exact acelaşi rezultat. Faptul de a aplica a doua oară acelaşi test influenţează rezultatul, de aceea putem vorbi despre oscilaţii nesemnificative ale rezultatelor la folosirea testului pentru examinarea aceleiaşi persoane de către un alt examinator. Subiectivitatea poate surveni în trei faze diferite ale derulării testării psihologice, motiv pentru care vorbim de tot atâtea feluri de obiectivitate: obiectivitatea de realizare, obiectivitatea de evaluare şi obiectivitatea de interpretare. În primul caz, modul de administrare a testului de către un examinator nu ar trebui să ducă la obţinerea unor rezultate semnificativ diferite la aceeaşi grupă de subiecţi faţă de cele obţinute de un alt examinator. În mod firesc, condiţiile de administrare a testului trebuie să fie cât mai asemănătoare, dacă se poate chiar identice, iar consemnul stabilit în caietul testului trebuie să fie respectat cu stricteţe.

Obiectivitatea de evaluare este mai greu de atins decât cea de realizare, iar cea de interpretare este mai greu de atins decât cea de evaluare.

Testele sunt diferenţiate în ceea ce priveşte dificultatea asigurării obiectivităţii sau concordanţei interpersonale. Asigurarea este mai simplă la testele creion-hârtie care posedă, în cea mai mare parte, reguli clare de producere a răspunsului corect, de corectare şi de evaluare şi mult mai grea la testele proiective, unde pregătirea psihologului pentru test

este de durată şi unde intervine un număr mai mare de criterii de evaluare şi de interpretare. Metoda statistică folosită pentru calcularea indicelui de concordanţă interpersonală este cea a corelaţiilor simple şi multiple. Calcularea obiectivităţii poate fi considerată o componentă a unei proceduri mai vaste, aceea a stabilirii fidelităţii unui test.

6.2.2. *Fidelitatea testului*

Această exigenţă faţă de scorul unui test se referă la precizia cu care un test măsoară o anumită caracteristică, indiferent dacă respectiva caracteristică este sau nu cea pe care pretinde că o măsoară (identitate care, o dată asigurată, indică validitatea testului).

Noţiunea utilizată în literatura engleză de specialitate pentru a desemna acest criteriu de valabilitate a unui test este *reliability*, termen ştiinţific de largă circulaţie. Literatura germană de specialitate foloseşte termenul *Die Reliabilität* pentru a exprima fidelitatea.

Noţiunea de fidelitate presupune că scorurile la un test trebuie să fie reproductibile (posibilitatea de a obţine aceleaşi rezultate prin repetarea măsurării, pentru aceleaşi persoane, în aceleaşi condiţii, cu teste măsurând aceeaşi trăsătură, în ocazii diferite).

Cunoaşterea mărimii fidelităţii este o condiţie necesară pentru o utilizare corectă a datelor, care au fost obţinute cu ajutorul unui instrument diferenţial psihologic. Pentru a stabili dacă un test măsoară cu o anumită siguranţă, se fac raportări la fluctuaţii intraindividuale şi interindividuale.

În spiritul concepţiei teoriei clasice a testului, se poate spune că diferenţa dintre scorul observat sau măsurat al unui test şi scorul său „adevărat" trebuie să fie cât mai mică. Exprimată sub formă grafică, această corespondenţă poate fi redată după cum urmează.

Raportul dintre cele două varianţe, a datelor adevărate şi a celor observate, defineşte fidelitatea unui test. Ca atare, fidelitatea are expresia:

$$\frac{s_T^2}{s_X^2},$$

unde cu s_T^2 se notează varianţa scorului „adevărat", iar cu s_X^2 se notează varianţa scorului măsurat.

Minimumul şi maximumul teoretic al acestui raport sunt 0,00 şi, respectiv, 1,00. Acest raport poartă numele de *coeficient de fidelitate*.

Scorurile obţinute de subiecţi în urma evaluării răspunsurilor date la solicitările testelor psihologice sunt failibile, adică supuse posibilităţii de a fi, în parte sau în totalitate, eronate şi, în consecinţă, suspectate de a nu exprima adevăratul potenţial. Am precizat în consideraţiile anterioare care pot fi sursele unor astfel de erori. Faţă de măsurătorile fizice, cele psihologice sunt mult mai expuse unor oscilaţii valorice datorate erorilor aleatorii sau celor sistematice. Aceste oscilaţii fac obiectul celor mai frecvente imputări pe care reprezentanţii ştiinţelor exacte le aduc măsurătorilor realizate de psihologi. Ca urmare a respectivei stări de fapt, constructorii de teste au preocupări speciale pentru a diminua ponderea erorii în măsurătorile lor şi, dacă acestea se produc totuşi, strădania constă în estimarea lor cât mai precisă.

Preocupările privind precizia rezultatelor testelor psihologice au două direcţii de orientare:

- una se referă la gradul de consistenţă a rezultatelor (gradul relaţiei dintre scorurile obţinute în condiţii diferite de efectuare a testelor);

- a doua cercetează cauzele diferenţelor observate între scoruri (factorii care generează scorurile inconsistente şi importanţa efectelor acestora).

Teoria clasică a testelor furnizează suportul teoretic pentru cercetarea preciziei, adică postulatele, definiţiile şi relaţiile care trebuie în mod necesar stabilite. Conceptele de *scor adevărat* şi *eroare* sunt fundamentale pentru teoria fidelităţii. Scorul adevărat, numit şi scor „platonic" de către F.M. Lord, M.R. Novick şi J.P. Sutcliffe, este acel scor obţinut la un test care măsoară fără eroare, dar această situaţie este ipotetică, deoarece, în domeniul empiric, în majoritatea cazurilor avem un anumit gradient de eroare. Din acest motiv, rezultatul unui test este un scor observat sau un scor failibil.

Conform celui de-al doilea postulat, scorul adevărat este speranţa matematică a scorurilor observate, adică scorul mediu obţinut de către un subiect în situaţia pur teoretică de utilizare de un număr infinit de ori a instrumentului de măsură numit test psihologic.

Conceptul de *eroare* este la fel de uzitat în procesul fundamentării teoretice a măsurării. Eroarea poate fi pozitivă, caz în care asistăm la o supraestimare a scorului adevărat de către cel observat, sau negativă, pentru situaţia contrară, adică de subestimare a scorului adevărat de către cel observat. Eroarea considerată în teoria clasică a testului psihologic se datorează în întregime hazardului. Sursele de eroare sunt independente unele de altele, în sensul că ele nu se află niciodată în corelaţie. La un număr mare de măsurări, valorile pozitive şi negative ale erorii se anulează reciproc şi media erorii tinde spre 0.

Determinarea fidelităţii se bazează teoretic pe două modele, şi anume modelul testelor paralele şi modelul de eşantionare într-un univers, ambele furnizând aceleaşi concluzii în privinţa erorii de măsură. Modelul testelor paralele este cel mai cunoscut şi utilizat, datorită simplităţii sale. Modelul de eşantionare într-un domeniu sau într-un univers este mai puţin folosit, deoarece necesită aplicarea mai multor formule şi principii.

Abordarea problemei fidelităţii presupune precizarea noţiunii de *test paralel*. Într-un mod general, testele paralele sunt definite ca serii de operaţii măsurând aceeaşi trăsătură în acelaşi grad. Maurice Reuchlin defineşte astfel paralelismul testelor: „Testele paralele sunt probe care au itemii de aceeaşi dificultate şi de aceeaşi natură"[1].

Există mai multe forme de paralelism al testelor, acestea diferenţiindu-se în funcţie de exigenţele exprimate prin postulate.

Testele strict paralele sunt constituite din teste în care mediile şi varianţele scorurilor observate, ca şi varianţele erorilor sunt identice pentru fiecare test, fiecare subiect având acelaşi scor adevărat în toate formele. Pentru majoritatea testelor utilizate în practica psihodiagnostică, aceste exigenţe sunt lipsite de realism.

Testele paralele τ-echivalente admit diferenţe între varianţele de eroare la două teste şi varianţele scorurilor observate, dar mediile la cele două scoruri adevărate trebuie să fie identice.

Testele esenţial τ-echivalente admit diferenţe între varianţele şi mediile scorurilor observate. Scorul adevărat pe care îl obţine un subiect la testul f este egal cu scorul pe care îl obţine la testul g plus o constantă. Constanta se raportează la test, nu la subiecţi. Dacă V_{pf} exprimă scorul adevărat pentru subiectul p la versiunea f şi V_{pg}, scorul adevărat al subiectului p la versiunea g, atunci relaţia liniară dintre cele două scoruri adevărate este exprimată prin ecuaţia următoare:

1. Vezi *Grand dictionnaire de la psychologie*, Larousse, Paris, 1992, p. 543.

$$V_{pf} = V_{pg} + C_{fg},$$

unde C_{fg} este o constantă, care depinde doar de testele f şi g şi nu depinde de subiect.

Exigenţele formei congenerice a paralelităţii sunt îndeplinite atunci când se consideră că scorurile adevărate ale celor două versiuni sunt perfect corelate. Relaţia liniară dintre scorurile adevărate ale subiectului p pentru versiunile f şi g ale unui test se exprimă astfel:

$$V_{pf} = b_{fg} + V_{pg} + C_{fg},$$

unde constantele b_{fg} şi C_{fg} depind doar de teste, nu şi de subiecţi. Această formă este funcţională mai ales atunci când numărul de itemi variază de la un test la altul. Varianţele scorurilor observate f şi g nu diferă numai din cauza eterogenităţii varianţei de eroare (ca în versiunea τ-echivalentă), ci şi din cauza eterogenităţii scorurilor adevărate. În anul 1971, Joreskog a arătat că testele paralele şi testele τ-echivalente pot fi considerate cazuri particulare ale formelor congenerice.

Corelaţia dintre două teste paralele este egală cu raportul dintre varianţa scorurilor adevărate şi varianţa unuia sau altuia dintre teste:

$$r_{12} = r_{11} = \frac{s_V^2}{s_1^2 \text{ sau } s_2^2}$$

În general, $r_{12} = r_{22} = r_{XX'}$.

Am stabilit anterior formula $s_X^2 = s_V^2 + s_E^2$. Dacă împărţim fiecare dintre elementele acestei formule cu aceeaşi cantitate (cu varianţa scorului observat, de exemplu), se obţine:

$$\frac{s_X^2}{s_X^2} = \frac{s_V^2}{s_X^2} + \frac{s_E^2}{s_X^2} = 1,$$

în care $\frac{s_V^2}{s_X^2}$ este proporţia varianţei observate provenind din varianţa adevărată, iar $\frac{s_E^2}{s_X^2}$ este proporţia varianţei observate provenind din varianţa erorii.

Precizia poate fi definită ca fiind un raport între varianţe: $r_{XV'} = \frac{s_V^2}{s_X^2}$ în care prin $r_{XX'}$ se notează coeficientul de fidelitate. Formula se poate rescrie în felul următor: $r_{XX'} = 1 - \frac{s_E^2}{s_X^2}$, în care $r_{XX'}$ reprezintă coeficientul de fidelitate. El corespunde unui raport de varianţe şi ia valori între 0 şi 1, deci nu poate fi decât pozitiv. Indicele de fidelitate este dat de corelaţia dintre scorul adevărat şi scorul observat. Formula sa este $r_{XV} = \sqrt{r_{XX'}}$. Acest indice reprezintă precizia cu care itemii permit măsurarea unei caracteristici anume.

Din formula erorii se pot realiza următoarele deducţii: $s_E^2 = s_X^2 - s_V^2$; dar:

$$r_{XX'} = \frac{s_V^2}{s_X^2}.$$

Drept urmare, $s_E^2 = s_X^2 - s_X^2 r_{XX'} = s_X^2(1 - r_{XX'})$, de unde:

$$s_E = \sqrt{s_X^2(1 - r_{XX'})} = s_X\sqrt{1 - r_{XX'}}.$$

S_E se numeşte eroarea standard a măsurării. În consecinţă, spunem despre o măsură că este fidelă sau precisă atunci când are cât mai puţine erori posibile sau când s_E este cât mai mică posibil. F.M. Lord şi M.R. Novick au demonstrat că procedura standard servind la estimarea coeficientului de fidelitate şi a varianţei erorii unui test presupune cel puţin două măsurători (scoruri) pentru fiecare respondent şi că precizia unor astfel de estimaţii se va mări o dată cu numărul observaţiilor asupra persoanei măsurate.

Considerând că erorile, după o infinitate de aplicări, se distribuie după modelul repartiţiei Gauss-Laplace, se poate spune, conform unor proporţii clasice rezultate din considerente statistice, că există 68% şanse ca valoarea adevărată să se afle în intervalul $X \pm s_E$, 95% şanse ca valoarea adevărată să se afle în intervalul $X \pm 2s_E$ şi 99,958% ca valoarea adevărată să se afle în intervalul $X \pm 3s_E$.

Dacă acceptăm constituirea unei mulţimi de teste paralele, plecând de la eşantioane de acelaşi volum de itemi extraşi la întâmplare dintr-o populaţie importantă de itemi posibili, atunci media fiecărui eşantion de itemi, astfel constituit, prezintă variaţii şi, conform rezultatelor din statistică, abaterea standard a unor asemenea variaţii va fi egală cu abaterea standard a populaţiei de itemi împărţită la rădăcina pătrată a numărului de itemi ai fiecărui eşantion, adică la eroarea standard de măsură a mediei. Având în vedere faptul că fidelitatea se pronunţă în primul rând asupra preciziei măsurării, un rol important în teorie îl joacă eroarea standard de măsură. Aceasta se mai numeşte şi eroarea tip de măsură.

Când se utilizează un scor observat sau failibil (vulnerabil la eroare), eroarea standard a măsurii dă o indicaţie asupra intervalului probabilist de încredere în care se situează scorul adevărat. Fără această informare asupra preciziei estimării, un scor este practic inutilizabil. Ca atare, media, varianţa şi eroarea standard de măsură sunt cei mai importanţi indici statistici în psihometrie. Cuantificarea fidelităţii unui test are ca raţiune obţinerea estimării erorii standard de măsură.

American Psychological Association a elaborat, în 1985, un document intitulat *Standards for Educational and Psychological Testing*, în care recomandă constructorilor de teste să comunice eroarea tip de măsură pentru diferite niveluri ale scorurilor.

Aşadar, pentru constatarea paralelismului în diverse forme, se precizează varianţa scorurilor observate şi a erorii, corelaţia scorurilor adevărate şi exclusivitatea hazardului pentru varianţa erorii. Datorându-se hazardului, erorile unui test prezintă o corelaţie nulă cu erorile unui alt test.

David Bartram (1994, p. 68) consideră că se poate ilustra adecvat constituirea testelor paralele în modul expus în continuare. Dacă testul este un ansamblu de itemi şi avem, de exemplu, un test de 100 de itemi, conceput pentru a măsura aceeaşi trăsătură, şi îl împărţim în două teste de 50 de itemi sau în patru teste de 25 de itemi, sau în zece teste de câte zece itemi, atunci subansamblurile respective de itemi se pot numi teste paralele dacă satisfac următoarele condiţii:

- toate subansamblurile de itemi trebuie să aibă aceeaşi medie, iar media trebuie să fie egală cu mediile scorurilor adevărate;
- ele trebuie să aibă aceeaşi varianţă, iar această variantă trebuie să fie egală cu varianţa scorurilor adevărate;
- indicele de corelaţie dintre oricare pereche posibilă de itemi trebuie să fie de aceeaşi valoare;
- subansamblurile de itemi trebuie să prezinte, toate, aceeaşi corelaţie cu orice altă variabilă care nu face parte din respectivul ansamblu de itemi.

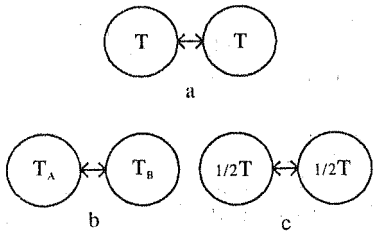
Acest mod de constituire a testelor paralele, prin divizarea şi subdivizarea testului, este unul ilustrativ pentru primul model de determinare a fidelităţii.

Al doilea model, prin eşantionarea unui univers de itemi, îl constituie amplificarea cantitativă şi calitativă de proporţii a primului model şi este mai apropiat de procedurile moderne de construire a testelor. Crearea unor teste paralele după acest model implică existenţa prealabilă a unor vaste ansambluri de itemi, denumite în termeni de specialitate *universuri de itemi*. Ideea de constituire a unor colecţii de itemi sau de paradigme de solicitare pentru teste nu este de dată recentă, dar constituirea efectivă şi gestionarea acestora nu a fost posibilă decât prin utilizarea tehnologiilor informatice moderne.

Metode de verificare a fidelităţii testelor. Metodele puse în practică pentru verificarea fidelităţii testelor sunt:

- metoda test-retest;
- metoda testelor paralele;
- metoda înjumătăţirii testului;
- metodele Kuder-Richardson;
- metodele mixte.

În spiritul teoriei clasice a testului, primele trei forme de fidelitate se pot vizualiza în modul următor:



unde *a* reprezintă fidelitatea test-retest, *b* fidelitatea test paralel, iar *c* fidelitatea prin înjumătăţirea testului.

Enumerarea acestor metode poate crea confuzie în corelaţie cu afirmaţiile anterioare, care indică posibilitatea realizării unor condensări, prin considerarea unui test paralel cu el însuşi şi considerarea fiecărei jumătăţi a testului ca fiind paralelă cu cealaltă. Explicaţiile din partea de prezentare a fiecărei metode vor aduce precizări suficiente.

Existenţa mai multor metode face să putem vorbi şi de mai multe interpretări ale coeficientului de fidelitate.

Există patru modalităţi de interpretare a coeficientului de fidelitate:

1) ca o corelaţie între scorurile observate şi scorurile adevărate;
2) ca un standard de comparaţie a preciziei unui test cu precizia altor teste de aceeaşi natură;
3) ca procentaj al respondenţilor care schimbă rangul şi poziţia;
4) ca un indice al gradului de eroare, în scorurile individuale.

Metoda test-retest. Maniera cea mai cunoscută de verificare a fidelităţii unui test este de a-l administra aceluiaşi eşantion de subiecţi, în două reprize diferite. Aceasta constituie metoda test-retest, a cărei problematică este discutată în continuare. Se

urmăreşte, în primul rând, stabilitatea în timp a rezultatelor succesive, introducându-se coeficientul de stabilitate.

Descompunerea varianţei totale, după Gulliksen, este dată de formula:

$$s_X^2 = s_V^2 + s_{EM}^2 + s_{EA}^2 + s_{EH}^2 + s_{ES}^2 + s_{VF}^2,$$

cu semnificaţiile prezentate deja (vezi *supra*, p. 125).

Nu există corelaţie între componentele varianţei totale. Eroarea totală, în cazul utilizării acestei metode, este egală cu suma ultimilor patru termeni ai ecuaţiei. Varianţa adevărată o cuprinde pe aceea care apare la două administrări (s_V^2) şi pe aceea atribuită memoriei (s_{VEM}^2).

David Magnusson face, de asemenea, consideraţii de diferenţiere a scorurilor obţinute. Varianţa care este tratată ca varianţă a erorii la calcularea corelaţiei între rezultatele celor două administrări ale testelor rezultă din următoarea expresie:

$$s_e^2 = s_{e(adm)}^2 + s_{e(g)}^2 + s_{e(subj)}^2 + s_{T(fl)}^2$$

Din această formulă rezultă că varianţa adevărată se obţine atât din s_T^2 (varianţa adevărată care apare la ambele prezentări ale probei), cât şi din $s_{e(m)}^2$ (varianţa erorii, care este dependentă de efectul amintirii). Coeficientul de fidelitate este:

$$r_{tt} = 1 - \frac{s_{e(adm)}^2 + s_{e(g)}^2 + s_{e(subj)}^2 + s_{T(fl)}^2}{s_t^2},$$

în care:

r_{tt} = coeficientul de fidelitate;

s_t^2 = varianţa totală a testului;

$s_{e(adm)}^2$ = varianţa erorii care este atribuită diversităţii modurilor de administrare a testului;

$s_{e(g)}^2$ = varianţa erorii care este atribuită efectului explicaţiilor date pentru efectuarea probei;

$s_{e(subj)}^2$ = varianţa erorii care este atribuită obiectivităţii deficitare;

$s_{T(fl)}^2$ = varianţa adevărată, care exprimă fluctuaţia valorilor adevărate.

Sunt necesare câteva explicaţii la formulele utilizate de Harold Gulliksen şi David Magnusson, în scopul de a evita o serie de neclarităţi. Această modalitate de a descompune varianţa observată a unui test constituie un aport benefic la încercările de a stabili sursele erorilor în scorurile unui test. Este o perspectivă care particularizează ceea ce era considerat globalizant până atunci. Această tendinţă este prezentă şi în cadrul unor teste statistice, precum λ^2 (aprofundarea analizelor prin fuzionarea efectivelor unor modalităţi şi prin analiza modalităţilor două câte două) şi ANOVA (prin testele particularizante Schäffe, Duncan, Bonferoni). Important este să se stabilească procedurile de detectare a acestor surse de eroare.

Influenţele ce pot interveni asupra rezultatului examinării prin intermediul unui test psihologic pot fi atribuite şi modului său de administrare. În cazul în care afectarea

randamentului respondenţilor se produce la un anumit punct temporal, fără să apară sistematic la repetare, aceasta va provoca o anumită varianţă a erorii, care, la calcularea fidelităţii, este considerată ca atare. În schimb, dacă una sau alta dintre influenţe, de exemplu efectul de interacţiune între examinator şi examinat, intervine sistematic (prin faptul că acelaşi examinator apare în ambele momente temporale ale aplicării testului), din această situaţie poate rezulta o varianţă aparent adevărată (care era atribuită efectului de reamintire). Acest fel de varianţă a erorii (de exemplu, eroarea datorată efectului amintirii) nu va intra în varianţa erorii de ansamblu, în circumstanţele date, la calcularea coeficientului de fidelitate.

Este evident că o persoană nu posedă constant o anumită capacitate, modificările survenind o dată cu trecerea anilor sau datorită diferitelor conjuncturi. H.J. Eysenck consideră că nu trebuie să ne concentrăm doar asupra aprecierii fidelităţii unui test, ci şi asupra fidelităţii persoanei. Această evidenţă motivează considerarea unei variante pertinente a valorii adevărate, în construcţiile teoretice privind fidelitatea. Dacă varianţa, care depinde de fluctuaţia valorii adevărate, este considerată varianţă a erorii, ea va contribui la subestimarea fidelităţii testului. Dacă, pe de altă parte, varianţa care ia naştere pe baza efectului de memorare (de reamintire a unor elemente constitutive ale solicitării unui test efectuat anterior) este tratată ca varianţă adevărată, se va produce supraestimarea fidelităţii testului. Prima posibilitate amintită creşte o dată cu mărirea intervalele între momentele temporale de administrare a testelor, iar a doua scade în aceste condiţii. Deoarece ambele variante trebuie luate în considerare, ca exprimări nedorite ale varianţelor, la calcularea fidelităţii intervalul temporal între aplicările succesive ale testului trebuie să fie ales în aşa fel încât efectul combinat să fie pe cât de mic posibil.

Câţiva teoreticieni ai testului sunt de părere că fluctuaţia în privinţa valorii adevărate a indivizilor între rezultatele temporale ale testului prezintă o varianţă care trebuie să fie luată în considerare la evaluarea fidelităţii unui test ca varianţă a erorii. Se poate distinge, în acest caz, între:

a) capacitatea unui test de a măsura gradul de reliefare a unei caracteristici;
b) stabilitatea respectivei caracteristici.

Este firesc să se considere că o măsurare este pe deplin sigură, chiar dacă se obţin, cu acelaşi instrument de măsurare, rezultate diferite pentru randamentul unei persoane la distanţe relativ mari în timp.

Există trei postulate de la care se pleacă în calcularea coeficientului de stabilitate:

- caracteristica măsurată este stabilă în timp;
- nu există efect de practică (postulat care poate fi criticat dacă intervalul este scurt, deoarece în acest caz respondenţii pot să-şi amintească răspunsurile, ceea ce poate mări corelaţia dintre rezultate);
- nu trebuie să existe instructaje diferenţiale între cele două administrări ale testului.

Astfel, între pretest şi posttest, corelaţia reflectă învăţarea diferenţiată (instructajul, experienţa), şi nu stabilitatea, deoarece respondenţii au achiziţionat între timp cunoştinţe suplimentare. În acest caz, estimarea preciziei este slabă. Dacă intervalul între două teste este prea lung, poate apărea învăţarea diferenţială, iar dacă este prea scurt, intră în joc efectele practicii. Eşantionul de solicitări nu afectează coeficientul de stabilitate, deoarece se utilizează acelaşi eşantion de itemi în cele două ocazii.

Aplicarea acestei metode implică mai multe dificultăţi. Astfel, efectul practic sau de memorie poate influenţa respondenţii la a doua administrare a testului, în aşa fel încât erorile înregistrate la cele două administrări au tendinţa de a fi în corelaţie. Mai mult, modificările de abilitate a respondenţilor contribuie, de asemenea, la eroare. Este greu să se precizeze dacă eroarea va avea drept efect mărirea sau diminuarea corelaţiei între rezultate. După părerea lui H. Gulliksen, corelaţia va fi mărită mai ales dacă intervalul de timp între două treceri va fi scurt, iar factorul „oboseală" intervine în mică măsură. În acest caz, corelaţia obţinută este o supraestimare a coeficientului de precizie. Dacă intervalul între două treceri este lung, memoria intervine mai puţin şi modificările în abilitatea respondenţilor devin mai importante, având drept efect reducerea corelaţiei. Dacă se consideră că fluctuaţia scorurilor adevărate este eroare, fidelitatea este subestimată, iar când efectul memoriei este considerat varianţă adevărată, fidelitatea este supraestimată. Varianţa care se poate atribui fluctuaţiei în scorurile adevărate se măreşte o dată cu durata intervalului între două administrări, pe când efectul memoriei o diminuează. Intervalul ideal trebuie să permită reducerea, pe cât posibil, a efectului combinat al acestor doi factori. Metoda test-retest este adecvată atunci când se doreşte o corijare în vederea atenuării efectelor prezentate.

Corelaţia rezultată după două administrări succesive ale aceluiaşi test se numeşte *indice de fidelitate test-retest*. Termenii de *indice* şi *coeficient* sunt similari. Acelaşi test, la două administrări succesive, constituie echivalentul a două teste paralele, deoarece aceiaşi itemi sunt utilizaţi în timpul ambelor administrări. Aşadar, practic, se măsoară stabilitatea temporală a aceluiaşi test, motiv pentru care coeficientul de corelaţie obţinut se mai numeşte *coeficient de stabilitate*. Durata de timp între cele două aplicări trebuie să fie semnificativă, recomandabil de ordinul lunilor, deoarece reaplicările imediate ale unui test conţin în rezultat mai multe influenţe favorizante. Metoda test-retest pune în relaţie rezultatele la acelaşi test în momente temporale diferite. Se constată, în general, diminuarea coeficientului de stabilitate o dată cu trecerea timpului şi se recomandă raportarea timpului scurs de la efectuarea ultimelor calcule asupra fidelităţii prin acest procedeu. Dacă perioada de timp scursă este de ordinul anilor, atunci ne putem aştepta la diminuarea coeficientului de stabilitate de la 0,90 (o corelaţie de nivel bun) până la 0,60 (o corelaţie de nivel modest, puţin satisfăcătoare).

R.B. Cattell distinge între o fidelitate test-retest de termen scurt şi una de termen lung, fiecare dintre ele furnizând informaţii diferite. Cea de termen scurt (două săptămâni şi chiar mai puţin) este denumită şi *coeficient de încredere*, indicând gradul de încredere ce se poate acorda instrumentului utilizat. În acest interval de timp, nu se poate vorbi cu adevărat despre transformări semnificative ale trăsăturii măsurate. R.B. Cattell păstrează termenul de *stabilitate* coeficientului test-retest care se realizează pentru un interval temporal important, susţinând că prin scăderea coeficientului de stabilitate din cel de încredere se poate obţine partea varianţei datorate fluctuaţiilor reale ale trăsăturii.

În cazul în care perioada de timp pentru aplicarea metodei test-retest este scurtă, prima efectuare a testului poate influenţa mărimea scorului la a doua efectuare, mai ales la testele de randament.

Se consideră că în stabilirea acestui tip de fidelitate există tendinţa de supraestimare, prin manifestarea varianţei reziduale specifice itemilor, care reprezintă varianţa sistematică specifică a acestora.

Metoda testelor paralele. Când testele paralele se constituie prin extragerea unor eşantioane dintr-o populaţie compusă din itemi de aceeaşi natură, coeficientul de corelaţie indică gradul de certitudine cu care poate fi măsurată o anumită trăsătură, aşa cum aceasta se prezintă prin populaţia de itemi. Astfel, mai multe eşantioane aleatorii, care conţin acelaşi număr de itemi, pot să constituie teste paralele şi permit obţinerea mai multor coeficienţi de corelaţie. Media acestor coeficienţi este cea mai bună estimare a coeficientului de corelaţie aşteptat şi verosimil pentru testele paralele de aceeaşi lungime. Numărul de itemi şi nivelul de omogenitate a populaţiei de itemi, adică gradul în care se măsoară aceeaşi variabilă, sunt cei doi factori determinanţi ai acestei medii. Coeficientul care indică gradul relaţiei între testele paralele, formate din eşantioane extrase aleatoriu, este un coeficient de consistenţă internă. În practică totuşi, testele nu satisfac exigenţele testelor strict paralele, deşi în ansamblu ele măsoară acelaşi scor adevărat, iar corelaţia între teste este pozitivă. Coeficientul de fidelitate, rezultat din corelaţia între testele paralele, va fi afectat printr-o varianţă proprie a fiecărui test. Chiar dacă este vorba despre varianţa adevărată, ea este considerată eroare. *Coeficientul de echivalenţă* este coeficientul reprezentat prin corelaţia între teste şi formele paralele, iar varianţa specifică are drept efect subestimarea adevăratului coeficient de fidelitate.

În mediul constructorilor de teste, există practica creării de forme paralele ale aceluiaşi test. Chestionarul de personalitate 16 PF, creat de R.B. Cattell, are până în prezent patru forme (A, B, C, D). Fiecare probă este compusă din itemi diferiţi, însă îşi propune acelaşi scop psihodiagnostic. În practica psihodiagnostică, se procedează în modul următor: dacă prima dată s-a aplicat forma A, retestul se face cu forma B, corelaţia dintre cele două forme numindu-se *fidelitate a formelor paralele*, iar coeficientul obţinut – *coeficient de echivalenţă*. Acest coeficient de fidelitate are, de regulă, o valoare mai mică decât a coeficientului de stabilitate. Fidelitatea este considerată atunci ca fiind un raport între rezultatele ambelor măsurători. Varianţa totală a unuia dintre cele două teste poate fi exprimată astfel:

$$s_t^2 = s_T^2 + s_{e(m)}^2 + s_{T(aqu)}^2 + s_{e(adm)}^2 + s_{e(g)}^2 + s_{e(subj)}^2 + s_{T(fl)}^2$$

Termenul care intervine în plus faţă de ecuaţia precedentă, aparţinându-i lui D. Magnusson, este $s_{T(aqu)}$, care, la estimarea fidelităţii, reprezintă varianţa adevărată considerată o varianţă a erorii la corelarea între testele paralele. Apariţia acestei varianţe de echivalenţă presupune existenţa unei componente adevărate specifice la un test, care nu este prezentă la celălalt test. Datorită considerării $s_{T(aqu)}$ ca fiind varianţă a erorii la calcularea coeficientului de echivalenţă, se produce o diminuare a acestui coeficient, deci o subestimare implicită a coeficientului de precizie. Formula coeficientului de echivalenţă devine:

$$r_{tt} = 1 - \frac{s_{T(aqu)}^2 + s_{e(adm)}^2 + s_{e(g)}^2 + s_{e(subj)}^2 + s_{T(fl)}^2}{s_t^2}$$

Semnificaţiile termenilor sunt cele date anterior. Apar aceleaşi probleme în legătură cu distanţa dintre perioadele de timp între aplicările testelor şi stabilirea efectului minim de fluctuaţie în valoarea adevărată. Cei care iau în considerare fluctuaţia valorii adevărate între aplicările unor teste, la estimarea fidelităţii, ca fiind varianţă relevată a erorii văd în coeficientul testului paralel cea mai bună exprimare pentru fidelitatea unui test.

Aceeaşi formulă o prezintă Gulliksen, în felul următor:

$$r_{XX'} = 1 - \frac{s_{VEQ}^2 + s_{EA}^2 + s_{EH}^2 + s_{ES}^2 + s_{VF}^2}{s_X^2}$$ (pentru teste paralele).

Metoda combinată. Reprezintă o altă metodă pentru estimarea fidelităţii constând în combinarea metodei test-retest şi a metodei testelor paralele. Procedura utilizată presupune administrarea unei forme de test pentru un grup de respondenţi, urmată de administrarea unei a doua forme de test pentru acelaşi grup şi readministrarea, după o perioadă de timp, a primei forme.

Exemplificăm cu următoarele date fictive. Să presupunem că am examinat un grup de 15 subiecţi cu forma T_1 a unui test, urmată, la aceeaşi dată, de forma echivalentă T_2 şi, peste doi ani, din nou cu forma T_1. Scorurile la cele trei administrări, însoţite de calculele necesare, sunt prezentate în tabelul următor. S-a notat cu X_1 scorul obţinut la prima administrare a testului T_1, cu X_2 scorul obţinut la a doua administrare a testului T_1 şi cu Y scorul obţinut la administrarea testului T_2.

Tabelul 5

Date necesare calculării coeficientului de fidelitate

Nr.crt.	X_1	Y	X_2	X_1^2	Y_2	X_1^2	X_1Y	X_1X_2	YX_2
1	32	30	33	1,024	0,900	1,089	0,960	1,056	0,990
2	34	35	36	1,156	1,225	1,296	1,190	1,224	1,260
3	35	35	34	1,225	1,225	1,156	1,225	1,190	1,190
4	30	34	35	0,900	1,156	1,225	1,020	1,050	1,190
5	33	35	34	1,089	1,225	1,156	1,155	1,122	1,190
6	37	37	38	1,369	1,369	1,444	1,369	1,406	1,406
7	40	42	41	1,600	1,764	1,681	1,680	1,640	1,722
8	41	41	42	1,681	1,681	1,764	1,681	1,722	1,722
9	40	42	41	1,600	1,764	1,681	1,680	1,640	1,722
10	36	38	39	1,296	1,444	1,521	1,368	1,404	1,482
11	38	37	39	1,444	1,369	1,521	1,406	1,482	1,443
12	39	40	38	1,521	1,600	1,444	1,560	1,482	1,520
13	35	36	37	1,225	1,296	1,369	1,260	1,295	1,332
14	37	38	38	1,369	1,444	1,444	1,406	1,406	1,444
15	38	39	40	1,444	1,521	1,600	1,482	1,520	1,560
Σ	545	559	565	19,943	20,983	21,391	20,442	20,639	21,173

Se obţine $\overline{X}_1 = 36{,}33$, $\overline{Y} = 37{,}26$, $\overline{X}_2 = 37{,}66$, $s_{X_1} = 3{,}10$, $s_Y = 3{,}24$, $s_{X_2} = 2{,}79$.

Din formula de calcul pentru corelaţie prin momentul produselor:

$$r_{xy} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right] \times \left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

se obţin rezultatele următoare: $r_{X_1Y} = 0{,}897$, $r_{X_1X_2} = 0{,}945$ şi $r_{YX_2} = 0{,}913$, care estimează coeficienţii de fidelitate discutaţi.

Când este dificilă stabilirea fidelităţii prin metoda test-retest, este foarte util să se ia în considerare două forme paralele ale aceluiaşi test. În acest caz, ne putem întreba până la ce punct scorurile obţinute la o formă de test sunt în concordanţă cu scorurile obţinute la altă formă de test.

În practică, se încearcă să se contrabalanseze efectele datorate ordinii de efectuare a testelor, procedându-se în modul următor: unei jumătăţi din grupul de subiecţi i se administrează întâi forma A şi apoi forma B a testului, administrarea testelor făcându-se în ordine inversă pentru cealaltă jumătate a grupului de subiecţi.

Postulatul de la care se pleacă în calcularea coeficientului de echivalenţă este următorul: formele testului sunt perfect echivalente referitor la conţinut, format, lungime, grad de dificultate, cu aceeaşi medie şi aceeaşi abatere standard. Inconsistenţa poate fi atribuită diferenţelor în eşantionarea itemilor. Principala sursă de varianţă în precizie provine din diferenţele dintre itemi la cele două forme paralele ale testului.

Metoda înjumătăţirii testului. În multe situaţii ale practicii psihodiagnostice, fie nu este posibil să se găsească nici un test paralel, fie nu se poate aştepta o perioadă de timp, în luni sau ani, pentru a aplica acelaşi test, existând doar posibilitatea administrării unei singure probe, o singură dată. Itemii celor două jumătăţi pot fi trataţi ca aparţinând la două teste paralele. Astfel, dintr-un test de 100 de itemi se pot constitui două teste de câte 50 de itemi. Coeficientul obţinut se aseamănă cu cel de echivalenţă, dar testul nu conţine în acest caz decât jumătate din cantitatea de itemi folosiţi la aplicarea unei forme test-retest sau la aplicarea ulterioară a unui test echivalent. În literatura engleză de specialitate, această metodă este cunoscută sub numele de *split-half*.

Există şi opinia potrivit căreia prin înjumătăţirea unui test nu se construiesc, de fapt, două teste paralele, ci se creează măsurători paralele, după care ambele jumătăţi sunt considerate teste paralele.

Pentru a realiza împărţirea testului în două, teoretic există mai multe posibilităţi. Practica verificării testelor şi metodologia psihologică au evidenţiat următoarele forme:

- *Înjumătăţirea în funcţie de itemii pari şi impari.* Oferim un scurt exemplu în această privinţă:

Tabelul 6

Date necesare calculării coeficientului de fidelitate prin metoda înjumătăţirii

Subiect	Itemi											
	1	2	3	4	5	6	7	8	9	10	S_1	S_2
DN	1	1	1	1	1	0	1	1	1	1	5	4
TR	0	1	1	0	0	1	1	0	1	0	3	2
MO	0	1	1	1	1	1	1	0	1	0	4	3
RS	1	1	1	0	0	0	1	1	0	1	3	3
CE	1	0	1	1	1	1	0	1	1	1	4	4
LI	1	0	0	1	0	0	1	1	0	1	2	3
FP	0	1	0	1	1	0	1	0	1	0	3	2
GA	1	1	1	1	0	0	1	1	0	1	3	4
DV	1	1	1	1	1	0	0	0	1	0	4	2

unde S_1 reprezintă suma itemilor impari, iar S_2 suma itemilor pari.

Această modalitate de înjumătăţire nu poate fi aplicată testelor în care exigenţa de rapiditate se află în prim-plan. Astfel, ar rezulta deformări considerabile ale coeficienţilor de fidelitate. Metoda poate fi aleasă atunci când este posibilă o singură prezentare a testului (de exemplu, itemii 1, 3, 5, 7, 9, pe de o parte, şi 2, 4, 6, 8, 10, pe de altă parte).

- *Înjumătăţirea la întâmplare (aleatorie).* În acest caz, se consideră că se pot evita erorile sistematice. Este rar aplicată, doar la sarcinile de dificultate egală ale unui test de nivel. Alegerea itemilor pentru a constitui un total parţial al unei jumătăţi este aleatorie, putându-se realiza o mulţime de combinaţii posibile.
- *Înjumătăţirea pe baza analizei datelor.* Reprezintă un procedeu modern, rezultat din construcţia sistematică a testelor paralele. Se formează, în acest caz, perechi de sarcini de dificultate aproximativ egală şi de discriminare asemănătoare. Din aceste perechi se extrage la întâmplare câte una pentru fiecare jumătate a testului. La acest tip de înjumătăţire se postulează că toate sarcinile au fost realizate de toţi subiecţii, că este vorba despre un pur test de nivel. Oricum trebuie să existe siguranţa că toţi subiecţii au avut suficient timp pentru a răspunde la toţi itemii înainte de a împărţi în două testul original. Este necesar ca testele să fie omogene într-un grad rezonabil şi de o lungime suficientă pentru ca prin coeficientul $r_{xx'}$ să se obţină o bună reprezentare a preciziei, dar cu condiţia să nu se facă o corecţie pentru atenuare.
  Calculele se pot face astfel: a) se calculează valoarea punctajului pentru fiecare jumătate şi b) cele două punctaje sunt corelate.
- *Înjumătăţirea după timpul necesar testării.* Se practică atunci când la un test predomină exigenţa de rapiditate. Dacă un test are ca timp de efectuare zece minute, se înregistrează rezultatele din primele cinci minute şi din ultimele cinci minute, care sunt folosite apoi pentru corelaţie.
  Unui număr considerabil de teste colective şi individuale, li se introduce o limită de timp, limită care poate fi considerată un factor determinant pentru rezultatele subiecţilor. Existenţa limitei de timp presupune că doar o parte dintre subiecţi pot aborda toate sarcinile. Un subiect care lucrează încet va obţine un număr redus de rezolvări corecte în comparaţie cu situaţia în care ar fi avut la dispoziţie un timp nelimitat. Testele al căror rezultat este influenţat de limita de timp sunt numite *teste de rapiditate* (*speed tests*). Ele se deosebesc de aşa-numitele *niveautests* (*power tests*), pentru care nu se prevede nici o limită de timp. Limita de timp poate fi apreciată diferit, în sensul că joacă un rol mai mare sau mai mic pentru rezultatul final.
  Un test pur de rapiditate constă din sarcini atât de uşoare încât pot fi rezolvate de majoritatea persoanelor examinate. Rezultatul testului nu este determinat de capacitatea individului de a putea să rezolve sarcini, ci de rapiditatea cu care acesta poate să abordeze sarcinile foarte simple introduse în test. Fiecare subiect va rezolva, în acest caz, toate sarcinile până la o anumită limită, care prezintă o măsură a rapidităţii. De exemplu, în proba *Discurile Walther*, în care se solicită transferarea unor discuri uşoare şi de dimensiuni mici din lăcaşurile unui plăci în cele ale altei plăci, toţi subiecţii pot realiza toate solicitările. Este însă important ca acest lucru să se facă într-un timp cât mai scurt. Ceea ce contează este timpul, deoarece subiectul nu poate realiza greşit o anumită sarcină. Având în vedere faptul că fiecare subiect poate termina un număr de sarcini prestabilit în timpi diferiţi, procedeul de înjumătăţire nu se poate realiza decât măsurând timpul pe care l-a realizat la parcurgerea a jumătate

dintre solicitări. Fidelitatea unui test pur de rapiditate este calculată prin corelaţia între două măsurări ale rapidităţii, care sunt obţinute independent una de alta, lucru ce se realizează fie prin metoda repetării testului, fie prin metoda testelor paralele. Este clar că metoda obişnuită a înjumătăţirii sarcinilor nu poate fi aplicată în estimarea fidelităţii măsurătorilor care se realizează cu testele de rapiditate.

La un alt tip de test pur de rapiditate este obţinută o măsură a rapidităţii prin numărul de sarcini prelucrate într-o perioadă de timp prestabilită. De astă dată, toate sarcinile parcurse (şi nu neapărat toate sarcinile cuprinse în test) sunt rezolvate corect. În această situaţie, se recurge la procedura de comparare a randamentelor rezultate în urma înjumătăţirii timpului prevăzut pentru întregul test (prima jumătate şi a doua jumătate). La testele creion-hârtie, după scurgerea primei jumătăţi a timpului de testare, examinatorul atenţionează examinaţii să marcheze locul unde au ajuns cu răspunsurile. După întrerupere, subiecţii fie efectuează sarcinile care urmează, fără întrerupere, fie încep să lucreze la o probă de la început. La probele aparative sau computerizate, se înregistrează randamentul obţinut prin mijloace automate. Practic, se produce separarea răspunsurilor primilor 50% dintre itemi de următorii 50% (sau a răspunsurilor la componentele de solicitare), rezultată firesc din evaluarea eficacităţii respondenţilor la cele două părţi ale testului, prin calcularea scorurilor respectivelor jumătăţi. Testele cu dificultate progresivă a itemilor nu sunt recomandate pentru acest tip de împărţire. Pe baza timpului de înjumătăţire, sunt obţinute două valori brute de acelaşi grad de independenţă, ca la înjumătăţirea sarcinilor.

Într-un test pur de nivel, tuturor subiecţilor li se oferă posibilitatea de a parcurge toate sarcinile, dar e probabil că vor exista şi răspunsuri incorecte. Răspunsurile se înregistrează într-o matrice de punctaje având ca elemente valori de 1 şi 0 (1 – răspuns corect, 0 – răspuns incorect). Variaţia distribuţiei valorilor obţinute într-un test de nivel va depinde de capacitatea individului de a rezolva corect sarcinile.

În situaţia în care s-a ajuns la constituirea jumătăţilor testului într-unul dintre modurile descrise anterior, este estimată precizia testului, prin tehnica împărţirii în două a acestuia, sub forma unui coeficient de echivalenţă. Se calculează coeficientul de corelaţie pentru estimarea nivelului concordanţei între rezultatele jumătăţilor prezentate. Acest coeficient de corelaţie se poate prezenta în chip de coeficient de fidelitate al uneia dintre cele două părţi ale testului care stau la baza calculului. Se acceptă faptul că jumătăţile sunt paralele şi se calculează coeficienţii de fidelitate pentru întregul test, utilizând una dintre formulele de corecţie. Pentru a satisface postulatele unui coeficient de echivalenţă, jumătăţile trebuie să fie independente şi egale în privinţa conţinutului lor, a gradului de dificultate, a mediei şi a abaterii standard.

După ce subiecţii efectuează testul, se calculează două scoruri pentru fiecare subiect, altfel spus câte unul pentru fiecare jumătate. Deoarece nu se operează decât jumătate dintre itemii utilizaţi în celelalte metode, se face apel la o serie de formule de corecţie, pe care le prezentăm în următorul tabel.

Tabelul 7

Formule de corecţie pentru coeficienţii de fidelitate

Autorul	Formula de corectare	Observaţii
W. Kristof	$r_{tt} = \frac{2}{n-1} + \frac{N-3}{N-1} \times \frac{4s_1s_2r_{12}}{s_1^2 + s_2^2 + 2s_1s_2r_{12}}$	Condiţii : $s_1 = s_2$ N = numărul de subiecţi
L.J. Cronbach	$r_{tt} = \frac{4(s_1^2 - s_1s_{(1-2)}r_{1(1-2)})}{4s_1^2 + s_{(1-2)}^2 - 4s_1s_{(1-2)}r_{1(1-2)}}$	$r_{1(1-2)}$ – corelaţia valorilor brute ale primei jumătăţi a testului cu diferenţele valorilor brute din ambele jumătăţi ale testului
L. Guttman	$r_{tt} = 2 \times \left(1 - \frac{s_1^2 + s_2^2}{s_X^2}\right)$	formă echivalentă a formulei Rulon
P.J. Rulon	$r_{tt} = 1 - \frac{s_{(1-2)}^2}{s_X^2}$	s_X – abaterea standard a întregului test $s_{(1-2)}$ – abaterea standard a diferenţelor $(X_1 - X_2)$, valorile brute ale primei jumătăţi minus valorile brute ale celei de-a doua jumătăţi
J.C. Flanagen	$r_{tt} = \frac{4s_1s_2r_{12}}{s_1^1 + s_2^2 + 2s_1s_2r_{12}}$	s_1 – abaterea standard a valorilor brute din prima jumătate a testului s_2 – abaterea standard a valorilor brute din a doua jumătate a testului
Spearman-Brown	$r_{tt} = \frac{2 \times r_{12}}{1 + r_{12}}$	postulatul că $s_1 = s_2$ nu este întotdeauna respectat

Cea mai uzitată dintre aceste formule este cea care a fost introdusă de Ch. Spearman şi W. Brown. Formula generală pentru varianta de corecţie Spearman-Brown este :

$$r_{nn'} = \frac{nr_{tt'}}{1 + (n-1)r_{tt'}},$$

unde :

- n este factorul prin care este mărită lungimea testului (noua lungime = lungimea actuală × n), formula din tabel fiind obţinută pentru $n = 2$;
- $r_{tt'}$ reprezintă precizia testului original ;
- $r_{nn'}$ este estimarea preciziei testului de n ori mai lung.

Această formulă generală permite să se afle până la ce punct trebuie lungit un test pentru a obţine gradul de precizie dorit. În cazul aplicării formulei obişnuite, estimăm practic fidelitatea testului în ansamblu.

Formula Spearman-Brown postulează că variabilitatea celor două jumătăţi ale testului este egală. Pentru cazul în care cineva nu poate să postuleze varianţe egale pentru părţile

testului, P.J. Rulon a luat în considerare distribuţia diferenţelor de scor între jumătăţi pentru a estima fidelitatea testului în întregime, utilizând ecuaţia următoare:

$$r_{XX'} = 1 - \frac{s_D^2}{s_X^2},$$

în care s_D^2 este varianţa diferenţelor dintre cele două scoruri.

Varianta Guttman de corecţie permite obţinerea unei estimări a preciziei, în funcţie de jumătăţi, prezentând fiecare o variabilitate diferită:

$$r_{XX'} = 2\left(1 - \frac{s_A^2 + s_B^2}{s_X^2}\right),$$

în care s_A^2 şi s_B^2 sunt varianţele jumătăţilor A şi B ale testului, iar s_X^2 este varianţa testului luat în întregime. Graţie acestei formule, nu mai este necesar să se calculeze corelaţia între cele două jumătăţi.

Când jumătăţile au varianţe asemănătoare, formulele lui Rulon şi Guttman dau o estimare a fidelităţii egală aceleia obţinute prin formula Spearman-Brown. Când varianţele diferă, formula lui Spearman-Brown dă totdeauna o estimare mai mare decât celelalte două.

Coeficientul α şi coeficienţii Kuder-Richardson. Poate cea mai cunoscută modalitate de a exprima simplificat consistenţa internă este dată de formula α a lui Cronbach:

$$\alpha = \frac{k}{(k-1)} \times \left[1 - \frac{\sum s_i^2}{s_X^2}\right],$$

unde:

- k = numărul de itemi;
- $\sum s_i^2$ = suma varianţei itemilor;
- s_X^2 = varianţa testului.

Teoria clasică a testelor strict paralele este conformă cu următoarele postulate:

1) itemii măsoară un factor (sau o trăsătură) şi numai unul;
2) corelaţiile interitemi sunt egale;
3) itemii au varianţe egale;
4) itemii sunt de dificultate identică.

Formula α a lui Cronbach presupune ca măcar primele trei condiţii să fie verificate.

J.F. Kuder şi M.V. Richardson dezvoltă un sistem de ecuaţii pentru aprecierea fidelităţii, care se dovedesc mai practice în utilizare. Aceste ecuaţii, în denumire prescurtată, sunt: K-R 2, K-R 8, K-R 14, K-R 20 şi K-R 21. Ecuaţiile Kuder-Richardson prezintă două forme: una în care se utilizează varianţa itemilor exprimaţi polihotomic şi alta pentru itemii cu răspuns dihotomic. Cazurile de utilizare a formulelor (vezi Bartram, 1994, p. 82) sunt următoarele:

- în cazul în care nici o supoziţie nu este verificată (K-R 2);
- în cazul în care doar supoziţia 1 este verificată (K-R 8);
- în cazul în care primele două supoziţii sunt verificate (K-R 14);
- în cazul în care primele trei supoziţii sunt verificate (K-R 20);
- în cazul în care toate cele patru supoziţii sunt verificate (K-R 21).

Pe plan formal, coeficientul α al lui Cronbach este echivalentul ecuaţiei K-R 20. Forma simplificată a ecuaţiei Kuder-Richardson 20 este următoarea:

$$r_{XX'} = \frac{k}{(k-1)} \times \left[1 - \frac{\sum p_i \times q_i}{s_X^2}\right],$$

în care p_i este indicele de dificultate al itemului i şi q_i este egal cu $1 - p_i$.

Formula Kuder-Richardson 21 a fost dezvoltată pentru a estima fidelitatea în cazul în care se postulează că itemii au acelaşi coeficient de dificultate:

$$KR_{21} = r_{XX'} = \frac{n}{n-1} \times \left[\frac{s_X^2 - n\bar{p} \times \bar{q}}{s_X^2}\right],$$

unde cu $\bar{p}$ se notează media coeficienţilor de dificultate (care estimează valoarea comună postulată a coeficienţilor de dificultate), iar $\bar{q} = 1 - \bar{p}$.

În cazul în care una sau mai multe dintre supoziţiile prezentate anterior nu se verifică, atunci coeficientul obţinut subestimează consistenţa internă.

Prezentăm un exemplu fictiv, din care să rezulte etapele ce trebuie urmate în aplicarea metodei Kuder-Richardson 20:

Tabelul 8

Date necesare în calcularea coeficientului de fidelitate Kuder-Richardson 20

	ITEMI														
Subiect	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Total**
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	13
2	1	1	0	1	1	1	0	1	1	1	1	1	1	1	12
3	1	1	0	1	1	0	0	0	0	0	1	1	1	1	8
4	1	1	1	1	0	1	0	1	0	1	0	1	1	1	10
5	1	0	0	0	0	0	1	0	1	0	1	0	0	0	4
6	1	1	1	1	1	1	1	1	1	1	0	1	1	1	13
7	0	0	0	0	0	0	1	0	1	0	1	0	0	0	3
8	1	1	1	1	1	1	1	1	0	1	0	1	1	1	12
9	0	0	0	0	0	1	0	1	1	0	1	1	0	0	5
10	1	1	0	1	1	1	1	1	1	1	1	1	1	1	13
11	0	0	1	1	0	1	0	0	0	1	0	1	1	0	6
12	0	1	1	1	1	1	1	1	1	1	0	1	1	1	12
f_i	8	8	6	9	7	9	7	7	8	8	7	10	9	8	111
p_i	0,67	0,67	0,50	0,75	0,58	0,75	0,58	0,58	0,67	0,67	0,58	0,83	0,75	0,67	
q_i	0,33	0,33	0,50	0,25	0,42	0,25	0,42	0,42	0,33	0,33	0,42	0,17	0,25	0,33	
$p_i \times q_i$	0,24	0,24	0,24	0,23	0,25	0,23	0,25	0,25	0,24	0,24	0,24	0,21	0,23	0,24	2,98

Deci k = 12, $p_i \times q_i = 3{,}03$, $s_X^2 = 13{,}52$.

$$K\text{-}R\ 20 = \frac{12}{12-1}\times\left(1-\frac{3{,}03}{13{,}52}\right)=\frac{12}{11}\times 0{,}77 = 0{,}85$$

Varianţa adevărată este deci dependentă de valoarea covarianţei între itemii unui test, care este determinată de intercorelaţiile şi varianţele itemilor. Acest fapt permite să se avanseze ideea că *K-R* 20, indicele de consistenţă internă, depinde de corelaţiile între itemi, deci de gradul în care se referă la acelaşi conţinut, altfel spus gradul de omogenitate a itemilor.

După trecerea în revistă a diferitelor metode de calcul al coeficientului de fidelitate, se poate concluziona că nu există o identitate valorică a rezultatului dacă folosim metode diferite. Este de presupus că surse diferite de eroare influenţează diferenţiat fidelitatea unui test. Pot interveni situaţii specifice în privinţa explicaţiilor date pentru efectuarea testului, modul de administrare a acestuia şi modul de corectare. Un rol poate juca chiar şi lipsa de stabilitate a trăsăturii în sine.

Oferim, în cele ce urmează, câteva exemple de stabilire a unor coeficienţi de fidelitate (*apud* J.L. Cronbach, 1960):

Test	Eşantion	Procedură	Tip de fidelitate	Coeficient
Test de aptitudini şcolare	370 de elevi în ultimul an de liceu	Kuder--Richardson	echivalenţă	0,94 verbal 0,91 numeric 0,95 total
Test de aptitudini generale al lui Pintner, forma K	203 elevi de 12 ani	Par-impar	echivalenţă	0,86
Test scurt de selecţie	Aproximativ 230 de candidaţi la un liceu de asistenţi sanitari	Forme paralele	echivalenţă	De la 0,87 la 0,93 pentru 3 secţiuni
Test scurt de selecţie	72 de operatori maşini calcul într-o bancă	Retest după 2 ani	stabilitate	De la 0,72 la 0,84 pentru 3 secţiuni

6.2.3. *Validitatea testului psihologic*

6.2.3.1. Consideraţii generale

În sens larg, *validitatea* este proprietatea unui model (enunţ, concept, indicator specific al unui test statistic) de a corespunde destinaţiei pentru care a fost creat.

Acest concept trimite la relaţia dintre teorie sau orice element teoretic (model, ipoteză, concept, instrument, test) şi realitatea empirică pe care teoria consideră că o reprezintă. Noţiunea de *validitate* are o legătură strânsă cu aceea de *adevăr*. A spune despre un model că este valid înseamnă a afirma că entităţile pe care le utilizează (de exemplu, procese, reprezentări, motivaţii, stadii, complexe) au o anumită formă de „realitate" şi că enunţurile conţinute sunt adevărate (exemple de enunţuri: „Nivelul de motivaţie modifică rapiditatea proceselor de percepţie" sau „Complexul lui Oedip este un constitutiv al personalităţii tuturor indivizilor"). Utilizatorul unui model, în funcţie de cadrul

epistemologic unde se situează, este nevoit să respecte definiţia termenilor teoretici şi să fixeze în acelaşi timp condiţiile necesar de îndeplinit de către un model valid.

Noţiunea de *validitate* este centrală pentru psihologia ştiinţifică, adică psihologia care încearcă să-şi fundamenteze afirmaţiile prin intervenţia procedurilor experimentale, cărora li se adaugă un sistem statistic adecvat de verificare a rezultatelor. Nu este exagerat să se afirme că psihologii se împart în două categorii: cei care se preocupă de validitatea afirmaţiilor, teoriilor, modelelor, conceptelor şi instrumentelor folosite de ei şi cei care nu se preocupă de acestea, ci se mulţumesc cu aserţiuni (afirmaţii care nu sunt demonstrate).

În teoria testului psihologic, validitatea este un concept central, beneficiind în mod prioritar de atenţia tuturor celor preocupaţi de aplicarea şi, mai ales, construirea testelor. De asemenea, în domeniul mai general al măsurării psihologice, validitatea este considerată aspectul central al psihometriei.

În calitate de concept generic, validitatea încearcă să răspundă la întrebările următoare:

- Care este caracteristica măsurată de test?
- Testul măsoară ceea ce pretinde a măsura în prezent?
- Tipul de test utilizat convine cu adevărat pentru folosirea care îi este destinată?
- Valorile testului permit estimarea imediată a unui randament?
- Testul poate lămuri psihologic noţiunea implicată?
- Informaţiile furnizate de test sunt utile pentru a lua decizii?
- Ce interpretări se pot da scorurilor obţinute la test?
- Ce predicţii se pot face din scorurile obţinute la test?
- Ce proporţie de varianţă rezultă pentru variabila măsurată prin test?

Termenul de *validitate* nu este folosit exclusiv în teoria testului psihologic. În teoria percepţiei, se vorbeşte despre *validitate ecologică*. Noţiunea a fost creată de Egon Brunswik pentru a desemna gradul în care covariază stimulii distali şi proximali. El s-a gândit la operaţia de bază a percepţiei, aceea de a învăţa să estimăm adevăratul stimul distal (distal = fizic) pe baza valorilor variabile ale stimulului proximal care excită sistemele receptoare. Cu cât este mai mare validitatea ecologică a unui semn proximal pentru o proprietate a unui stimul distal, cu atât e mai probabil ca acesta să fie învăţat şi folosit. Noţiunea este utilizată în teoria percepţiei directe a lui J.J. Gibson.

Problematica validităţii unui procedeu diagnostic trebuie să fie privită în raport cu stabilirea măsurii în care un test îndeplineşte exigenţele fixate aprioric faţă de el. Aceste exigenţe nu trebuie să se refere totdeauna la localizarea unui individ folosind o anumită metrică. Un test poate să fie utilizat în întregime pragmatic în scopul realizării unei selecţii fără să se recurgă la capacităţile sau caracteristicile individuale ale subiecţilor. Validitatea poate fi privită prin această prismă, a deciziei pentru anumite scopuri, şi ne putem întreba pentru ce decizii este valid un test. Astfel, validitatea nu este o caracteristică generală a unui test, ci un test poate să fie valid pentru un scop sau altul.

Validitatea este calitatea testului psihologic de a măsura cu precizie o caracteristică pe care pretinde că o măsoară. În concepţia lui Serge Legendre (*apud* Bernier şi Pietrulewicz, 1997, p. 179), validitatea este „capacitatea unui instrument de a măsura realmente ceea ce trebuie să măsoare, după utilizarea pe care vrea să o facă". În spiritul teoriei clasice a testului psihologic, Gustav Lienert (1967, p. 255) considera că validitatea unui test dă gradul de precizie cu care acest test măsoară faptic acea caracteristică a personalităţii sau acel mod de comportament pe care el trebuie să-l măsoare ori pretinde să-l măsoare.

În concepţia lui Jürgen Guthke (Guthke, Böttcher şi Sprung, 1991, p. 248), validitatea sau valenţa diagnostică este „imaginea pe cât posibil de exactă şi adecvată a gradului de reliefare a unei structuri psihice". Un test este deplin valid atunci când rezultatele sale permit o concluzie nemijlocită şi fără echivoc asupra gradului de reliefare a caracteristicii psihologice cercetate.

Importanţa validităţii este punctată în teoria structurală a testului. În teoria clasică a testului, gradul de reliefare a unei caracteristici psihice este raportat la o populaţie de referinţă, în teoria probabilistă a testului reliefarea este o constantă parţială de grup, iar în teoria testului orientat spre un criteriu, ea ajută la realizarea unui scop particular (cum ar fi terapia şi învăţarea).

Preocupările sistematice privind validitatea testului psihologic s-au conturat la începutul secolului XX, prin contribuţiile lui Ch. Spearman şi F. Krueger, care au creat o procedură de corelaţie statistică pentru problema controlului validităţii, diverse caracteristici de randament, devenite mai târziu criterii de validitate, corelându-se între ele. Preocupările privind studiul validităţii au cunoscut transformări considerabile în ultimii 50 de ani, după apariţia lucrărilor teoretice de fundamentare a teoriei clasice a testului. Tendinţa care se conturează, în studiul validităţii, este aceea de a-i conferi o viziune unitară. American Psychological Association şi American Educational Research Association (cunoscute sub formele prescurtate de APA şi AERA) au publicat împreună, în anul 1985, *Standard for Educational and Psychological Testing*, în care se pronunţă pentru integrarea diferitelor tipuri de validitate. În documentul amintit, punct de referinţă pentru psihometricienii americani, se afirmă că validitatea este „un concept unitar care nu se referă la testul în sine, ci la inferenţele făcute plecând de la rezultatele acestuia" (*apud* Laveault şi Grégoire, 1997, pp. 189-190).

Pentru testele psihologice, condiţiile standard de validitate sunt definite, în general, în termeni de valoare predictivă, fiind exprimate prin intermediul unor corelaţii ale valorilor testului cu valori ale variabilelor-criteriu care presupun înglobarea a ceea ce testele sunt învestite să diagnosticheze (inteligenţa, nivelul de dezvoltare, trăsătura de personalitate, capacitatea într-o sarcină sau într-o profesiune). Este cazul cel mai des întâlnit, care nu epuizează conţinutul noţiunii de *validitate*.

Spre deosebire de validitate, *validarea* unui test psihologic este un ansamblu de proceduri cantitative şi calitative care au drept finalitate asigurarea validităţii. Este bine cunoscut faptul că noţiunile teoretice folosite în disciplinele socioumane, şi nu numai aici, sunt justificabile printr-o procedură în care ele sunt confruntate cu datele. În consecinţă, reguli metodologice riguroase fixează, pe baze raţionale, etapele pe care trebuie să le parcurgă o validare sistematică. În acest proces, elementele teoretice pot să fie ori respinse ca nonvalide, ori ameliorate, ori acceptate ca provizoriu valide (până la o confirmare mai amplă). Validarea este întotdeauna legată de o situaţie particulară, altfel spus, ea ar avea o finalitate diferită dacă ar fi obţinută în alte condiţii, folosind alte eşantioane sau alte metode de analiză. În nici un caz validitatea nu poate fi garantată pentru întregul ansamblu al situaţiilor sau contextelor în care poate fi folosit un test. Afirmaţia că un test este mai valid decât altul nu are sens, decât dacă cel dintâi se dovedeşte a fi valid într-o mai mare varietate de situaţii sau pentru un număr mai mare de scopuri.

6.2.3.2. Tipuri de validitate

Calitate complexă a testului psihologic, validitatea se stabileşte în funcţie de mai multe criterii, care precizează finalitatea rezervată testului. Ca atare, există atâtea tipuri de validitate câte categorii de obiective se cer asigurate de către test. În viziunea clasică, se diferenţiază obiectivele practice de cele teoretice. Testul poate constitui un factor determinant în decizie şi în acţiune, în situaţii foarte diverse : diagnosticul stării cunoştinţelor sau a competenţelor în vederea formării ori a consilierii, orientarea şi selecţia profesională, alocarea şi realocarea posturilor într-o organizaţie etc. Afirmaţiile validităţii trebuie specificate şi concretizate în funcţie de criteriile alese, de dependenţa faţă de eşantioane, de problematica diagnostică, de situaţia de decizie. Această stare de fapt, desemnată de H. Westmeyer ca validitate diagnostică, este formalizată în definiţia validităţii ca functor cu şase poziţii, desemnat prin termenul compus din variabilele < t, c, p, u, v, z >. În această exprimare vorbim despre o validitate a testului t, în relaţie cu criteriul c, la aplicarea pe clasa de persoane p, în condiţiile ambientale u, prin cercetătorul v, în domeniul temporal z (domeniu temporal în care indicatorul validităţii rămâne constant).

În condiţiile amintite, este posibil ca acelaşi test să aibă o validitate ridicată la o anumită categorie de subiecţi (sau la o anumită problematică) şi una scăzută la altă categorie de subiecţi (sau la altă problematică).

H. Wottawa (*apud* Guthke, Böttcher şi Sprung, 1991, pp. 248-249) consideră că din această situaţie decurg două consecinţe practice :

1. dezvoltarea de teste cu un domeniu limitat de valabilitate şi criterii mai precise de decizie pentru care este posibilă o precizare concretă de validitate ;
2. obligaţia practicienilor de a determina ei înşişi, prin propriile cercetări, dacă un procedeu recomandat în prevederile manualelor ca având o validitate bună (exprimată printr-un coeficient global de validitate) este valid şi într-o situaţie particulară de decizie.

Folosind o tipologie metodico-teoretică se poate realiza o separare grosieră între *validitatea logică* şi *validitatea empirică*. Prin validitate logică se desemnează un gen de validitate care apare rar în psihologie, formele subsumate acestei noţiuni (aşa-numita validitate totală sau validitate reprezentativă) referindu-se la cazuri strict determinate, cazuri ce nu cunosc limitări. Termenul de validitate logică se mai foloseşte în clasificările validităţii, motiv pentru care recomandăm o raportare la contextul clasificativ pentru înţelegerea sa nuanţată. Validitatea empirică este un termen generic pentru acele forme ale validităţii care fac posibile afirmaţii verosimile, condiţionate stocastic în privinţa cazurilor de referinţă. Această diferenţiere, făcută de R. Dieterich, L. Sprung şi H. Sprung în anii '80, este teoretic metodologică, şi mai puţin practic diagnostică.

D.T. Campbell distinge, în 1957, între validitatea internă şi cea externă, forme pe care le utilizează nu numai în teoria testelor, ci şi în metodologia psihologică generală.

Validitatea internă se referă la validitatea variabilelor urmărite într-o situaţie de cercetare, pentru lămurirea fenomenelor supuse analizei. Influenţa examinatorului sau influenţa instrumentului de măsură asupra rezultatelor măsurării sunt exemple de variabile care pot fi considerate tipice pentru asigurarea validităţii interne.

Problema care apare în cazul validităţii externe vizează stabilirea măsurii în care condiţiile de cercetare intern valide sunt reprezentative şi generalizabile pentru fenomenele

încă necercetate din aceeaşi clasă de fenomene, de exemplu, în ce măsură un test de inteligenţă reprezintă şi randamentul inteligenţei care este pretins în studiu sau în profesie. Erorile de selecţie a eşantioanelor, efectul Hawthorne, efectul placebo etc. prejudiciază validitatea externă.

O tipologie devenită tradiţională face diferenţa dintre validitatea de conţinut (*content validity*), validitatea de criteriu şi validitatea de construct. Această împărţire a validităţii a fost promovată de L. Sprung, H. Sprung, M. Nowakowska, L. Michel, W. Konrad ş.a., fiind recomandată şi de către APA.

O altă perspectivă de clasificare oferă luarea în considerare a tipului de concluzii care se pot trage din testele apreciate ca valide. Un tip de concluzii (extrapolare) evoluează de la comportamentul etalat la test la aşa-numitele trăsături latente, dintre care exemplificăm aptitudinile, caracteristici ale personalităţii, care condiţionează comportamentul la test, dar nu sunt accesibile nemijlocit observaţiei. Înţelegerea generală a validităţii este primordial legată de acest fel de concluzii. Validitatea de construct este un exemplu de asemenea concluzie. Alt gen de concluzii (extrapolări) sunt cele care se fac plecând de la comportamentul la test spre comportamentul exterior situaţiei de testare.

În practică, verificarea validităţii a fost realizată în special pentru acest ultim tip de concluzii, adică tipul orientat empirico-behavioristic (accesibil observaţiei directe). Este un tip de concluzie divizat, în continuare, într-o variantă la care este trasă o concluzie directă de la comportamentul la test la acelaşi fel de comportament exterior situaţiei de testare (concluzii de reprezentare) şi o variantă la care este acceptat un raport (o corelaţie) între comportamentul la test şi anumite aspecte ale comportamentului exterior de alt fel (criteriu) ; pe această bază este trasă aşa-numita concluzie corelativă. Pentru concluziile vizând comportamentul exterior de acelaşi fel, vorbim despre o validitate de reprezentare (care se împarte, la rândul ei, în validitate de conţinut şi validitate logică), iar pentru concluziile vizând comportamentul exterior de alt fel vorbim despre validitatea raportată la criterii sau validitatea empirică.

Prezentate sintetic, validităţile raportate la tipul de concluzii care se pot trage se prezintă astfel :

- concluzii de reprezentare (validitate de conţinut, cu variantele validitate aparentă şi validitate logică) ;
- concluzii corelaţionale (validitate referitoare la criteriu, validitate empirică) ;
- concluzii bazate pe factori latenţi (validitate de construct, respectiv reconstruct).

De asemenea, ne putem aştepta ca un test să constituie un eşantion reprezentativ al unui ansamblu de sarcini sau de conduite. În acest caz, testul va avea o *validitate de conţinut*.

Indiferent de toate preocupările de aplicare, ne putem întreba ce semnificaţie are măsurarea realizată prin intermediul unui test. Dacă această măsurare este ghidată, sprijinită, fundamentată printr-un ansamblu de propoziţii coerente şi adaptate, se va spune că un test are o bună *validitate teoretică*.

În primul rând, vom face consideraţii asupra primelor trei tipuri de validitate, care au fost general acceptate în rândul specialiştilor. Validitatea referitoare la criterii este separată în mod obişnuit în validitate predictivă şi validitate concurentă (vom face ulterior precizările necesare). Văzută din punct de vedere istoric şi practic, validitatea referitoare la criterii este cea mai importantă şi mai uzitată diviziune a clasificării validităţii. În definiţiile validităţii cuprinse în tratatele de psihodiagnostic, acest tip de

validitate este ales ca bază de referinţă, cu toate că dezbaterile asupra altor aspecte, cum ar fi validitatea de construct, au contribuit mult la lămurirea teoriei validităţii. Oricum, validitatea bazată pe criterii joacă un rol predominant. Vom trata cele trei ramuri principale ale validităţii, în ordinea facilităţii structurale, de la simplu la complex.

Validitatea de conţinut. În sens strict, un test psihologic posedă o validitate de conţinut (*content validity*, în terminologia engleză) când măsoară faptic ceea ce trebuie să măsoare după avizul anumitor persoane autorizate în acest sens. Validitatea de conţinut presupune acceptarea evidenţei că un test este expresia unui eşantion reprezentativ dintr-un ansamblu de itemi sau de sarcini (denumit *univers de itemi* sau *univers de sarcini*), considerat de către un consiliu de experţi ca fiind fundamentat cantitativ-calitativ şi desemnat ca indicat pentru măsurarea unei anumite caracteristici a persoanei. Într-o viziune restrânsă asupra acestui gen de validitate, referitoare mai ales la domeniul selecţiei profesionale, însuşi conţinutul testului prezintă criteriul optimal pentru caracteristica reliefată. Ca atare, dacă se urmăreşte cunoaşterea randamentului unui subiect în universul situaţiilor în care testul este eşantion, conţinutul testului constituie aspectul primordial ce trebuie luat în considerare, iar testul devine atunci universul în care itemii sunt eşantioane. Astfel, precizia răspunsurilor şi procesul de rezolvare a itemilor testului sunt de un mare interes. Evaluarea validităţii de conţinut se face în funcţie de gradul de pertinenţă al eşantionului de itemi. Caracteristicile care trebuie cuprinse, în privinţa validităţii de conţinut, sunt atât de evidente încât se vorbeşte despre ea ca despre o *validitate trivială*. Deoarece nu există un indice cantitativ de pertinenţă a eşantionului, evaluarea se face prin judecată şi prin apreciere raţională. Un test de cunoştinţe şcolare va fi valid în privinţa conţinutului dacă sarcinile sale prezintă alegeri reprezentative din materia de învăţământ. Cel mai ilustrativ exemplu pentru acest tip de validitate este proba la maşina de scris. O modalitate de a evalua validitatea unui test de abilităţi de lucru la birou va fi să stabilim gradul relaţiei între scorurile la test şi abilităţile cunoscute ale unui grup de lucrători la birou ale căror performanţe au fost evaluate în condiţii reale de activitate. Alte exemple ar fi testele de scriere corectă sau de rapiditate a reacţiilor motrice. La ultimul exemplu, există pericolul unei generalizări, care ar merge prea departe atunci când din rezultatele unui test de rapiditate a reacţiilor motrice s-ar trage concluzii asupra rapidităţii în general. Precizarea limitelor în aprecierea concluziilor unui test este obligatorie. În cazul validităţii logice, ca variantă a validităţii de conţinut, se pune accentul pe rigoare în eşantionare. Ea se prezintă, în general, sub forma unui tablou de specificaţii, care conţin delimitări clare ale subiecţilor investigaţi, determinări de obiective precise şi o selecţie cât mai logic posibilă a itemilor în funcţie de împărţirea coerentă a materiei de învăţare sau a domeniului vizat.

În viziune strictă, estimarea validităţii de conţinut a unui test se bazează pe o examinare detaliată a conţinuturilor itemilor din test. Evaluarea acestora se face analitic, prin examinarea atentă a fiecărui item în vederea stabilirii adecvării cu testul şi prin examinarea coeziunii globale a itemilor prin test. De exemplu, la un test de abilităţi matematice cu un grad respectabil de validitate de conţinut, ar trebui ca formularea itemilor să fie foarte clară pentru ca subiectul să înţeleagă ceea ce este întrebat. Mai mult, conţinuturile ar trebui să fie astfel gândite încât toate aspectele testate să fie reprezentate adecvat. Testul nu trebuie, de exemplu, să fie supraîncărcat cu itemi despre înmulţire în defavoarea itemilor despre adunare. Acest tip de validitate nu este considerat

esenţial, mai ales atunci când se pot stabili alte tipuri pertinente de validitate, apropiate destinaţiei psihodiagnostice.

Deoarece stabilirea unui asemenea tip de validitate depinde de experţi, precizăm, în problema apartenenţei experţilor la o categorie profesională, că, în cazul testelor de caracteristici psihice (caracter, dispoziţii), aceştia trebuie să fie psihologi care, în urma experienţei acumulate după o perioadă de lucru în echipă, să aibă pe cât posibil o viziune teoretică unitară asupra caracteristicilor persoanei aflate în discuţie. La testele orientate spre învăţare, spre terapie, precum şi la testele de capacităţi profesionale, experţii pot să fie psihologi, medici, profesori de diverse discipline, pedagogi, diferiţi formatori profesionali etc. Aceşti experţi trebuie să decidă dacă testul în întregime şi itemii individuali corespund unui univers de sarcini anterior delimitate, din care testul reprezintă doar un eşantion.

Cei care apreciază itemii nu pot fi consideraţi infailibili, pentru că nu decid apodictic într-un domeniu foarte vast, ci prezintă elemente de relativism. Ca atare, este necesar să analizăm calitatea judecăţii lor prin raportare la anumite repere. Sintetic, se pot rezuma trei repere:

1. Coerenţa internă a judecăţilor stabileşte persistenţa gradului de exigenţă a judecătorului de-a lungul efectuării actului apreciativ. Un judecător nu poate emite aprecieri foarte exigente doar pentru o parte dintre itemi, ci pentru ansamblul sarcinilor din testul supus analizei.
2. Varianţa judecăţilor: între judecăţile emise de diferiţi experţi nu trebuie să existe o prea mare diferenţă.
3. Concordanţa judecăţilor se realizează în cazul în care acestea au tendinţa de a ordona itemii în aceeaşi manieră, după gradul lor de congruenţă cu faţetele supuse măsurării.

Pentru precizarea validităţii de conţinut prin varianta sa denumită *validitate logică* (recunoaşterea testului de către experţi pe baza analizei logico-ştiinţifice), s-au propus în literatura de specialitate diferiţi indicatori de măsură. Lawshe a propus un coeficient de validitate de conţinut, CVR (*content validity ratio*), care are următoarea formulă:

$$CVR = \frac{N_e - \frac{N}{2}}{\frac{N}{2}},$$

unde N_e este numărul evaluatorilor (experţilor) care consideră testul, respectiv itemul, ca fiind reprezentativ, iar N este numărul total de evaluatori (experţi, decidenţi).

Precum un coeficient de corelaţie, coeficientul de validitate de conţinut variază între -1 şi +1. Cu cât valoarea este mai mare, cu atât mai valid este testul, respectiv itemul, în privinţa conţinutului.

Prin intermediul unui coeficient de concordanţă, R. Fricke pune în relaţie varianţa în interiorul decidenţilor cu varianţa maximal posibilă a decidenţilor. În 1980, K. Lindner definitivează tabelele de semnificaţie pentru precizarea concordanţei. Formula este următoarea (*apud* Guthke, Böttcher şi Sprung, 1991, p. 251):

$$Ü = 1 - \frac{s^2_{emp}}{s^2_{max}},$$

unde Ü este coeficientul de concordanţă (*Übereinstimmungskoeffizienten*), s^2_{emp} este varianţa în interiorul experţilor şi s^2_{max} este varianţa maximal posibilă între experţi.

Un procedeu deosebit de exigent pentru validitatea de conţinut presupune constituirea a două grupe de constructori de teste, care trebuie să descrie, independent una de alta, exclusiv pe baza informaţiei despre destinaţia psihodiagnostică, un *univers de itemi* şi care să dezvolte reguli generative pentru producerea de sarcini ale testului. Testele rezultate din această procedură trebuie să satisfacă exigenţele unor verificări statistice. Proceduri statistice special create sunt puse în funcţiune pentru a oferi o bază cantitativă acestor verificări, cele mai cunoscute fiind coeficientul lui Kendall, indicele lui Hambleton şi coeficientul K al lui Cohen.

Testele care sunt valide în privinţa conţinutului ar putea fi considerate instrumente de măsură pentru criteriu, nu teste ce sunt postulate ca valide. În acest caz, conceptul de *validitate de conţinut* posedă un anumit grad de neclaritate, de ambiguitate. Există o atitudine de contestare faţă de acest fel de validitate, de îndoială cu privire la utilitatea noţiunii în practica psihodiagnostică. Cea mai puternică obiecţie susţine că acestui concept nu îi este oferită nici o posibilitate de verificare empirică rezumativă. „Vocaţia acestui fel de validitate este adesea o simplă părere, mai ales când este vorba despre decizia unor specialişti", spune Angela Schorr (1993, p. 563). Un psihometrician german, P. Drenth, considera, în 1969, că acest fel de validitate se situează în imediata vecinătate a *validităţii de faţadă* sau a *validităţii aparente*. Ca atare, consideră autorul german, ar trebui să fim prudenţi când vorbim despre respectivul fel de validitate ca despre o *validitate logică*. Ceea ce este denumit *logic* este, în fapt, o afirmaţie recomandată de autorul testului ca fiind fondată pe judecată şi, în astfel de cazuri, se poate considera că autorul în cauză vrea să subscrie la confirmarea empirică a presupunerilor sale. Ca urmare a unor asemenea consideraţii, se recunosc două tipuri de validitate de conţinut, validitatea aparentă şi validitatea logică (fundamentată prin înţelegere, cercetare şi analiză psihologică).

Validitatea bazată pe criterii. Este precizată prin raportarea punctajului testului la valorile-criteriu, raportare care se face, de regulă, prin calcularea corelaţiilor. Calea spre dobândirea de către un test a validităţii bazate pe criterii este predominant empirică, motiv pentru care mai este denumită şi *validitate empirică*. Alegerea criteriului de validitate este un pas decisiv în procesul validării. Ca atare, trebuie să existe un interes permanent asupra modului în care a fost obţinut criteriul de validitate, deci asupra *validităţii* criteriului de validitate. După părerea lui G. Lienert, există pericolul unui regres al relaţiei de validitate, care poate să apară din precizarea criteriului de către evaluatori. Nivelul de competenţă al acestor evaluatori (de regulă psihologi) poate diminua şi chiar anihila acest pericol. Uneori, în operaţiunea de stabilire a criteriului sau criteriilor, mai ales la testele de aptitudini, sunt atraşi şi alţi experţi (profesori, formatori). Cu o astfel de procedură (acord, votări), criteriul poate fi văzut ca relativ, rezultat al unei convenţii, al unui acord. Rămâne la latitudinea fiecărui evaluator de test dacă va accepta sau va respinge un anumit criteriu, dacă va căuta altul mai bun. În fiecare caz, coeficientul de validitate calculat prezintă o informaţie foarte importantă despre test şi poate fi interpretat prin raportare la criteriul luat drept fundament.

Convenţia de adoptare a criteriului prin consiliere, ca orice convenţie, nu este ceva absolut şi poate fi supusă permanent schimbării. La multe proceduri de validare nu este

posibil, din motive tehnice, să se stabilească aprioric un criteriu satisfăcător, fiind acceptate variante de compromis, care sunt schimbate progresiv pentru soluţii mai bune. R.L. Thorndike deosebeşte, în această privinţă, între *criterii provizorii*, *intermediare* şi *finale* (*immediate*, *intermediate* şi *ultimate criteria*, în termenii psihologului american).

Criteriile provizorii utilizează primele informaţiile care se pot obţine despre subiect (de exemplu, în producţie, după ce un muncitor a fost selectat profesional, printr-un examen în care s-au folosit teste psihologice, primele impresii ale maistrului la cinci sau şase luni de la angajare).

Criteriile intermediare introduc mai multe informaţii, acestea fiind mai fundamentate. Se trece la o fază în care are loc o aproximare a realului. În cazul testului de selecţie profesională exemplificat anterior, pot să fie combinate decizia maistrului şi rezultatele unor examene periodice ale noilor angajaţi.

Criteriile finale utilizează cantitativ şi calitativ acele informaţii care pot fi considerate bine fundamentate pentru caracteristica diagnosticată. La modul ideal, criteriul final reprezintă adecvat caracteristica de prognozat. Toate criteriile utilizate în operaţii de validare vizează, de regulă, criteriul final de validitate.

Exprimat sub formă grafică, nivelul validităţii bazate pe criterii se poate prezenta astfel:

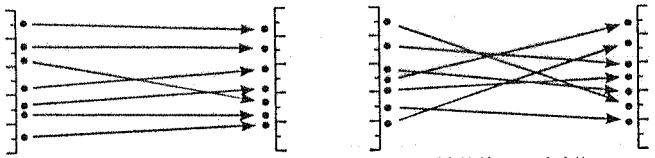


Figura 8. *Reprezentare spaţială pentru validitatea bazată pe criterii*

Se observă că o validitate înaltă reflectă corespondenţa aproape perfectă între valori de acelaşi grad, în timp ce o validitate slabă denotă o slabă asociere între valori de acelaşi grad.

Aceste precizări întăresc supoziţia potrivit căreia nu se poate vorbi de o singură validitate, ci de mai multe. Un test poate fi validat prin mai multe criterii şi, ca atare, poate avea mai mulţi coeficienţi de validitate, cu diverse valori. Dacă un test indică, după utilizarea unor criterii diverse, validităţi de niveluri valorice diferite, se poate afirma că testul posedă o *validitate diferenţială* (*differential validity*) referitoare la aceste criterii. Există cel puţin atâtea validităţi câte criterii sunt acceptate.

Un coeficient de validitate este determinat cu referire la un grup precizat de subiecţi. Ca urmare, un test poate să fie suficient de valid în cazul unui grup, dar nu poate fi pentru altul. L. Michel vorbeşte de validitate pe *subgrupe de prognozabilitate diferenţială*. Variabilele care separă fundamentat astfel de subgrupe au fost numite de D.R. Saunders *variabile moderatoare*.

Validitatea bazată de criterii este frecvent împărţită şi în funcţie de alte criterii de clasificare. Amintim două dintre acestea:

1. validitatea raportată la criterii interne şi validitatea raportată la criterii externe;
2. validitatea predictivă şi validitatea concurentă sau concordantă.

Cuplurile clasificative *validitate de prognoză/ validitate concurentă* şi *validitate externă/ validitate internă* nu se exclud unul pe altul, ci avem de-a face practic cu o intersectare a sferelor noţiunilor.

Validitatea raportată la un criteriu extern se stabileşte prin corelarea valorilor testului cu valorile unui *criteriu*, evaluat obiectiv ori prin valorile unei decizii argumentate obiectiv (de exemplu, admis sau respins). Validitatea unui instrument de testare este evaluată, în acest caz, prin determinarea relaţiei dintre scorurile la test şi un criteriu independent. Dacă un test pretinde că măsoară o anumită caracteristică psihică, este firesc ca testul să coreleze cu un criteriu care include aceeaşi trăsătură sau una foarte apropiată.

Validitatea predictivă. Dacă performanţa cu care se compară rezultatele testului urmează să fie stabilită ulterior examinării, validitatea este *predictivă*. Rezultatele testului sunt utilizate pentru a stabili predicţii cu privire la randamentul unui criteriu. Ca exemple de criterii folosite frecvent, avem media cumulată la absolvirea unor cursuri de formare sau nivelul productiv individual într-o activitate direct productivă. Acest tip de validitate se aplică atunci când se încearcă să se afle gradul în care scorurile obţinute la un test permit să se prezică randamentul la un criteriu, ceea ce constituie indicele de predicţie ca măsură a validităţii predictive a testului. Aspectul cel mai important al acestui tip de validitate este criteriul, deoarece scorul obţinut la un test nu este pertinent decât în măsura în care pretinde că prezice criteriul. Testele de aptitudini şcolare servesc la prezicerea şanselor de reuşită în instituţii de învăţământ, iar inventarele de interese permit distingerea câmpurilor de interese profesionale, inventarele de personalitate sunt utile pentru a distinge dezordinile psihologice care pot predispune un subiect să comită un anumit act generator de accidente productive.

Corelaţia între succesul prognozat şi cel realizat practic la un grup experimental este numită *validitatea predictivă a testului* (sau a bateriei de teste) *în relaţie cu un criteriu special*. Pentru determinarea corelaţiei respective se poate utiliza tehnica regresiei multiple. Această corelaţie s-ar putea numi *validitatea testului* sau *a bateriei de teste*, în funcţie de existenţa unui singur predictor sau mai multor predictori. Dacă pe baza unuia sau mai multor predictori şi a unui criteriu vrem să precizăm ecuaţia de regresie, este indicat, de exemplu, să se realizeze testul-predictor la o grupă de subiecţi concurenţi pentru admiterea la un curs de formare, dintre care doar o parte vor lucra în profesiunea în interesul căreia se efectuează selecţia. Testele sunt realizate înainte ca un candidat să înceapă activitatea. Rezultatele testelor sunt păstrate până când subiecţii au lucrat suficient de mult timp, deci până când rezultatele activităţii lor în respectiva profesiune pot fi exprimate cantitativ. Abia atunci este estimată ecuaţia de regresie pentru valorile-predictor şi valorile-criteriu.

Validitatea predictivă este pretinsă de un mare număr de teste, mai ales în controlul valabilităţii testelor în care prognoza este, într-adevăr, de o importanţă decisivă. Perioada de timp dintre evidenţierea rezultatelor testării şi a valorilor-criteriu poate să fie foarte diferită, de la câteva zile până la câţiva ani. S-ar putea calcula astfel diferiţi coeficienţi de validitate, corespunzători spaţiului temporal dintre administrarea testului şi producerea criteriului. În cazul validării predictive, scorul în sine la test nu are atâta importanţă cât gradul de predicţie ataşat la un randament sau la un comportament în contextul unui criteriu. Este important ca şi itemii să satisfacă o relaţie statistică cu un anumit criteriu, însă esenţial este ca rezultatul de ansamblu al testului să demonstreze o *validitate*

empirică, situaţie în care scorurile observate la test pot să permită prezicerea unui randament la un criteriu.

În numeroase cazuri, verificarea validităţii se face prin corelarea valorilor obţinute de subiecţi la testul care trebuie validat cu valorile obţinute la testele recunoscute anterior ca valide şi care diagnostichează aceeaşi caracteristică. Se vorbeşte în acest caz despre *validare internă* şi *validitate internă*. Stabilirea validităţii interne presupune, pe de o parte, că sunt utilizate teste care măsoară aceeaşi trăsătură şi, pe de altă parte, că aceste teste posedă un înalt coeficient de validitate sau sunt valide în privinţa conţinutului (este unul dintre sensurile validităţii interne). Un astfel de coeficient de validitate este denumit şi coeficient de validitate *secundar* (*second hand*) şi este, în majoritatea cazurilor, mai scăzut decât cel *principal*; există însă şi posibilitatea ca ambele teste să coreleze cu o a treia caracteristică psihică, necunoscută şi necontrolabilă.

Validitate şi prognoză. Variabilele-predictor sunt utilizate în scopul de a prognoza variabilele-criteriu. Variabilele-predictor sunt numite uneori şi variabile independente, iar variabilele-criteriu – variabile dependente. La prognozele statistice, variabilele independente sunt cele pe baza cărora sunt făcute estimări sau prognoze ale variabilelor dependente. Cele mai importante trei caracteristici ale variabilelor-predictor sunt următoarele:

- variabilele-predictor ca atare au o importanţă socială redusă;
- cheltuielile de măsurare sunt relativ reduse;
- ele pot să fie înregistrate, din punct de vedere temporal, înaintea variabilelor-criteriu.

Faptul că o variabilă este considerată ca fiind predictor sau criteriu este relativ, această calitate depinzând de contextul de cercetare. O variabilă poate servi într-o situaţie drept criteriu, iar în alta drept predictor. Când o variabilă este considerată criteriu, înseamnă că este apreciată ca având importanţă socială, adică examinatul este recompensat sau sancţionat pentru randamentul său. Exemplul cel mai concludent este cel al unui test folosit pentru măsurarea nivelului de randament la sfârşitul unui an şcolar, practică întâlnită în instituţiile de învăţământ americane. Astfel, testele de randament, care au fost utilizate la studenţii dintr-un an superior, pot să fie criterii pentru testele-predictor realizate la începutul studiului. În acest caz, testul de randament este un predictor pentru succesul în criteriul *randament la colegiu*.

Se vorbeşte despre *validitate temporală* în cazul testelor utilizate pentru prognoza criteriului într-o serie de activităţi speciale. Cu ajutorul ponderărilor de regresie se pot ordona variabilele-predictor la estimarea criteriilor. Pentru stabilirea ponderărilor, este necesar să se măsoare pe un singur eşantion de subiecţi atât variabilele-predictor, cât şi cele referitoare la criteriu. Ambele trebuie să fie disponibile pentru a putea calcula ponderile de regresie. Abia când acestea există la un eşantion, pot să fie aplicate şi la alte eşantioane, caracterizate de un singur predictor, şi trebuie să fie estimată valoarea verosimilă a criteriului. Pentru precizarea ponderii regresiei se apelează la două proceduri. Acestea se deosebesc prin ordinea temporală, iar predictorul şi criteriul sunt înregistraţi la grupa experimentală sau la cea de control. Una dintre metode duce spre aşa-numita *validitate predictivă*, cealaltă spre *validitatea concurentă*.

Interpretarea erorii unui coeficient de validitate internă se recomandă mai ales atunci când validarea se realizează doar prin raportare la valorile unui singur test, care este, pe

cât posibil, moderat valid. Aprecierea va trebui să se bazeze mai puţin pe coeficientul de corelaţie obţinut şi mai mult pe gradul validităţii logice. O situaţie mai avantajoasă apare atunci când validitatea internă se obţine prin raportarea la valorile mai multor teste individuale şi toţi coeficienţii de validitate rezultaţi sunt satisfăcători, conferind astfel o anumită încredere în coeficientul de validitate obţinut. Este posibil să se constituie un *coeficient de validitate multiplă*, prin intermediul căruia se studiază corelaţia testului nou introdus în raport cu ansamblul rezultatelor celorlalte teste. Un astfel de coeficient este *coeficientul de corelaţie multiplă*, R_{tc}.

Calcularea coeficientului de validitate multiplă, pentru trei sau mai multe teste, trebuie să se facă prin aşa-numiţii *coeficienţi beta*.

Se dă un exemplu în care un test nou este validat prin două teste confirmate.
Să presupunem că un test nou (testul 1), construit pentru a măsura abilităţile manuale, este corelat în scopul validării sale cu testul O'Connor (testul 2) şi cu testul discurilor Walther (testul 3), probe cunoscute, considerate valide. Rezultă următoarele corelaţii: $r_{12} = 0,7$ şi $r_{13} = 0,6$. Din intercorelaţia testelor cunoscute 2 şi 3 rezultă: $r_{23} = 0,3$. Coeficientul de validitate multiplă rezultat este următorul:

$$R_{tc} = \sqrt{\frac{0,7^2 + 0,6^2 - 2 \times 0,7 \times 0,6 \times 0,3}{1 - 0,3^2}} = 0,810$$

Coeficientul multiplu de validitate obţinut este mai mare decât fiecare dintre ambii coeficienţi de validitate internă. Este natural să fie mai mare, deoarece el include mai multe aspecte ale caracteristicii de analizat. S-ar putea să ne bazăm, prin urmare, pe acest fapt în alegerea testelor indicate pentru validarea internă.

Corelând *luarea deciziilor şi validitatea predictivă*, nu se încearcă niciodată să se prezică un criteriu pentru simpla plăcere de a o face. Predicţia vizează în mod esenţial luarea unei decizii şi, în acest sens, validitatea predictivă constituie un excelent instrument.

Când trebuie să se ia decizii referitoare la selecţie, la clasare, la plasare, se utilizează adesea scorurile individuale ale unor teste, furnizând informaţii pe care se poate sprijini o luare de decizie, şi alte informaţii relevante din diverse surse. Determinarea importanţei relative a scorurilor la teste, în raport cu informaţia provenind din alte surse, este, adesea, un demers ştiinţific dificil. F.G. Brown susţine că stabilirea validităţii predictive măreşte şansele ca o decizie luată plecând de la scorurile unui test să fie mai precisă, oricare ar fi gradul de exactitate al unei decizii luate plecând de la alte surse de informaţie.

Validitatea unui test este apreciată în funcţie de valoarea corelaţiei dintre scorurilor sale (scorurile obţinute pe un eşantion semnificativ de subiecţi şi valorile obţinute, pentru acelaşi eşantion, la criteriu) şi de gradul relevanţei criteriului, situaţie în care scorul-criteriu nu reprezintă în mod temeinic scopul vizat, corelaţia fiind, în realitate, un artefact matematic. Problema constă în luarea de decizii legate mai mult de grup decât de indivizi. Validitatea este întotdeauna estimată plecând de la datele grupului şi, logic, enunţurile cuprinzând validitatea unui test trimit întotdeauna la validitatea unui ansamblu de scoruri, şi nu la scorul unui respondent.

Criteriul. În planificarea construcţiei unui test, există un raport strâns între domeniul validităţii, pe de o parte, şi definirea unui criteriu adecvat de validitate, pe de altă parte. Alegerea unui criteriu reprezentativ de validitate constituie una dintre cele mai grele sarcini din dezvoltarea testelor, contribuind decisiv la succesul final al unui test. În 1944, H.A. Toops, reputat specialist american în psihometrie, autor al unor lucrări pe tema stabilirii criteriului, afirma în legătură cu acest subiect: „După posibilităţi, ar trebui să fie utilizat la fel de mult timp pentru definitivarea unui criteriu ca şi pentru construirea şi perfecţionarea testelor. Acestui domeniu important de cercetare i se dedică rar jumătate din timpul sau aşteptarea pe care el o pretinde şi o presupune" (*apud* Lienert, 1967, p. 265).

Se utilizează doar acele concepte ale validităţii la care sunt folosite variabile-criteriu şi decizii de valoare socială. *Utilitatea socială* trebuie să fie bine precizată pentru a nu se ajunge în dificultate metodologică. Variabilele de comportament pentru care societatea emite sancţiuni punitive sau premiale se pot recunoaşte uşor, fiind diferite de acele moduri de comportament care sunt lipsite de relevanţă din punct de vedere social.

În studiile realizate asupra validităţii predictive, criteriul este elementul cel mai important. Interesul pentru un test-predictor este totdeauna funcţie a relaţiei acestuia cu criteriul. Chiar atunci când un test are menirea de a măsura un randament, trebuie stabilit un standard de reuşită, iar acest standard este criteriul. În educaţie, se foloseşte aproape sistematic media cumulată drept criteriu. Nu înseamnă că nu pot fi folosite şi alte criterii, deoarece media cumulată nu este, în mod necesar, cel mai bun criteriu. Alte criterii pot fi aptitudinea de integrare, analiză, înţelegere, nivelul culturii generale, capacităţile de adaptare, manifestarea atitudinilor creative în faţa situaţiilor dificile. Se subînţelege că aceste criterii trebuie să se prezinte sub formă cuantificată. În cazul cel mai simplu se ia în considerare un singur criteriu, dar, în majoritatea situaţiilor, sunt necesare mai multe, fapt ce ridică o serie de probleme, în speţă complicarea calculelor. Atunci când criteriile sunt puternic legate unul de altul, combinaţia mai multora dintre ele nu aduce mare lucru în plus faţă de un singur criteriu, considerat din start ca fiind fundamental, deoarece măsoară acelaşi lucru cu celelalte. În cazul în care criteriile sunt slab legate între ele, nu este recomandat, cel puţin statistic, să fie combinate, deoarece ele măsoară lucruri diferite.

Dilema pusă în combinaţiile de criterii este desemnată prin conceptul de *dimensionalitate statistică*. E.E. Ghiselli relevă alte probleme care apar în această privinţă. Prima rezidă în faptul că un criteriu, considerat din start ca fiind fundamental, poate deveni inadecvat în timp, chiar dacă eşantionul şi contextul rămân aceleaşi. Avem în acest caz problema *dimensionalităţii dinamice*. A doua problemă este aceea a *dimensionalităţii individului* cuprinzând diferenţele individuale. Astfel, doi indivizi pot să realizeze aceeaşi activitate procedând în moduri foarte diferite şi ambii să fie cotaţi totuşi la fel de bine. Specialiştii se confruntă cu următoarea dilemă: a accepta cele două metode de realizare a unei activităţii ca fiind la fel de valide sau a privilegia una în detrimentul alteia.

Diviziunea criteriilor de validitate. O primă sistematizare a criteriilor de validitate a realizat în anul 1928 C.L. Hull, care a diferenţiat trei grupe de criterii, şi anume criteriul produs, criteriul-acţiune şi criteriul rezultat dintr-o impresie subiectivă.

O exemplificare pentru un *criteriu produs* este numărul produselor finite industriale realizate de către un muncitor într-o perioadă de timp delimitată. Un *criteriu-acţiune* ar putea fi rapiditatea (valoarea timpului de realizare) sau precizia cu care se efectuează

o anumită activitate. Pentru randamentele care nu pot fi evaluate obiectiv se poate opera cu *criteriul impresiei subiective*. Criteriul produs şi criteriul-acţiune pot fi desemnate cu noţiunea de *criterii de randament* şi contrapuse criteriului de apreciere sau de estimare subiectivă.

Alte optici de diviziune cu care s-ar putea opera iau în *considerare gradul complexităţii unui criteriu, gradul de obiectivitate* şi *gradul de cuantificabilitate*. Oferim în continuare, ca exemplu, sistemul de clasificare conceput de G. Lienert (1967, pp. 265-266). În toate cele trei puncte de vedere se prelevează semnificaţia practică a validităţii de criteriu.

1. a) Criterii simple de validitate.
   b) Criterii combinate de validitate:
      - criterii de validitate combinate simultan;
      - criterii de validitate combinate succesiv.
2. a) Criterii de validitate obiective.
   b) Criterii subiective de validitate.
3. a) Criterii calitative de validitate.
   b) Criterii cantitative de validitate şi criterii calitative de validitate gradate cantitativ.

Prin *criteriu simplu de validitate* se înţelege acel criteriu care corespunde unui domeniu relativ delimitat de validitate a unui test, de exemplu: venitul realizat prin practicarea profesiei; numărul de produse vândute într-o perioadă de timp pentru capacitatea profesională a unui vânzător; notele la examen pot fi considerate criteriu simplu şi subiectiv cantitativ pentru cunoştinţele de specialitate; aptitudinea sau inaptitudinea exprimată prin decizii competente pentru o anumită profesiune poate fi considerată criteriu simplu şi subiectiv cantitativ.

Criteriile combinate de validitate se compun din două sau mai multe criterii parţiale (simple), situaţie în care trebuie să fie calculate ponderi egale sau diferite pentru criteriile individuale parţiale. Criteriile parţiale pot să fie disponibile concomitent sau în succesiune temporală. Pentru succesul profesional, s-ar putea compune *criterii simultan combinate*, precum numărul de clienţi anual, venitul lunar rezultat din practicarea profesiei, alegerea în funcţii din cadrul asociaţiilor profesionale. În evaluarea aptitudinii de a conduce un mijloc de transport s-ar putea lua în considerare următoarele criterii parţiale: vechimea posesiei unui permis de conducere; numărul de accidente de circulaţie cauzate; numărul de încălcări ale regulilor de circulaţie consemnate în documente; mărimea distanţei parcurse de la primirea carnetului de conducere; frecvenţa reparaţiilor necesare la mijlocul de transport etc.

O *combinaţie succesivă de criterii parţiale* ar putea fi succesiunea temporală a datelor marcante ale maturizării infantile precoce: răspunsul la excitaţii, apucarea activă a obiectelor, ţinerea capului în poziţie verticală, şederea, mersul, vorbirea, păstrarea curăţeniei.

Criterii obiective pot fi considerate cele exprimate nemijlocit, sub formă de număr. De exemplu, numărul de încercări necesare pentru însuşirea deplină a unei deprinderi poate fi criteriu pentru aptitudinea de învăţare motorie, iar timpul necesar pentru traducerea unui anumit text într-o limbă străină poate fi considerat criteriu pentru gradul de stăpânire a respectivei limbi. Toate criteriile neclar obiective ar trebui să fie considerate subiective.

Separarea în *criterii cantitative* şi *calitative* se intersectează adeseori cu separarea în *criterii obiective şi subiective*, mai ales în cazul în care cele obiective sunt la fel de

frecvente ca şi cele cantitative, în timp ce criteriile subiective pot fi considerate atât cantitative, cât şi calitative. Calitativul este frecvent exprimat printr-o separare alternativă: da – nu; mai bun – mai rău; existent – inexistent.

Pentru nevoia de contact social, criterii obiective cantitative pot fi considerate apartenenţa la anumite asociaţii, contactul social frecvent cu vecinii, participările la întruniri festive cu rudele, prietenii, cunoştinţele, participarea la discuţii deschise etc. Pentru criteriile subiective şi concomitent cantitative sunt de amintit, ca exemple tipice, notele şcolare şi cele de la examene.

Cunoaşterea caracteristicilor criteriului implică necesitatea de a aprecia fundamentat un criteriu, deoarece construirea unor bune criterii este o operaţiune costisitoare, în timp şi bani. Se impun o serie de observaţii:

• *Criteriile nu există direct şi imediat.* Un criteriu nu se poate stabili imediat într-o activitate profesională sau într-o relaţie de predicţie, ci pretinde o anumită perioadă de timp pentru a putea fi exprimat. Un criteriu pentru integrarea profesională ar putea fi deosebita satisfacţie în activitatea pe care o desfăşoară subiectul şi mulţumirea cu veniturile materiale rezultate; un alt criteriu ar putea fi existenţa aplicării unei sancţiuni. Este important să se ştie dacă acestea se repetă sau nu, după trecerea unei perioade de timp considerate. Succesul într-o activitate sau apariţia unui anumit comportament presupune trecerea unei anumite perioade de timp. Nu putem să apreciem un student decât în momentul examenului. De asemenea, un alergător nu este apreciat decât după terminarea unei curse, nu după locul pe care-l ocupă la un anumit moment temporal al desfăşurării acesteia. Trebuie amintit faptul că stabilirea practică a unor criterii este o activitate costisitoare, cheltuielile pentru măsurarea mai multor criterii putând fi foarte mari în timp, efort, bani şi materiale.

• *Demersul statistic.* Găsirea unei exprimări cantitative pentru criterii este foarte importantă, indiferent dacă se face în domeniul activităţilor industriale, administrative sau educative. Problema este complexă şi dificilă, deoarece în privinţa criteriului apar variabile continue şi/sau discrete. Deoarece există mai multe criterii, atunci când se vorbeşte despre validitatea unui test psihologic, trebuie specificat criteriul cu care se corelează. Se poate spune că un test are atâtea validităţi câte criterii sunt definibile. În primul rând, se poate pune mai degrabă întrebarea „câte validităţi s-au stabilit până acum?", decât întrebarea „cât de înaltă este validitatea unui test?". În calcularea indicelui de validitate, se pot întâlni următoarele situaţii:

- un singur predictor – un singur criteriu;
- un singur predictor – mai multe criterii;
- mai mulţi predictori – un singur criteriu;
- mai mulţi predictori – mai multe criterii.

Un singur predictor – un singur criteriu. Cazul cel mai simplu este un singur criteriu – un singur predictor. În 1964, A.W. Astin stabileşte o distincţie între criteriu şi *măsură-criteriu*. Autorul amintit consideră criteriul ca fiind conceptul general ataşat la un randament reuşit (criteriu conceptual). Acest concept trebuie să fie exprimat în termeni operaţionali, după o măsură care poate permite determinarea validităţii unui test. De exemplu, reuşita la universitate este un criteriu conceptual, pe când media cumulată este o măsură-criteriu. Măsura-criteriu trebuie să posede anumite caracteristici, ce vor fi prezentate în ordinea importanţei lor.

Caracteristica cea mai importantă a unei măsuri-criteriu este pertinenţa, adică validitatea sa. Ea trebuie să reflecte şi să măsoare aspecte fundamentale ale criteriului conceptual. Astfel, media cumulată este o măsură-criteriu validă, deoarece ea reflectă şi măsoară corect aspectele importante ale reuşitei la universitate. Se evaluează pertinenţa unei măsuri-criteriu după o procedură raţională şi calitativă. Nu pare să existe un indice calitativ eficace pentru a exprima gradul de pertinenţă a unei măsuri-criteriu.

A doua caracteristică a măsurii-criteriu este fidelitatea sau constanţa. Aceasta presupune ca o măsură-criteriu să fie constantă în timp şi în diferite circumstanţe (sau să varieze într-un mod previzibil). O măsură-criteriu neconsistentă nu poate fi pusă în relaţie cu o alta variabilă, care are calitate de predictor, deoarece indicele acestei relaţii poate să ia atunci orice valoare, „la bunul-plac" al circumstanţelor, fără a semnifica niciodată nimic. Dacă se utilizează o măsură inconsistentă pentru a stabili productivitatea unui muncitor, în una dintre zilele sale „bune", productivitatea sa va fi ridicată şi se va supraestima valoarea acesteia, dar, dacă se pleacă de la o zi „rea", din contră, se riscă să se subestimeze.

Alegerea unei măsuri-criteriu este supusă la două constrângeri:

- factorii care induc o eroare de măsură trebuie să fie controlaţi, pentru că ei influenţează măsurile-criteriu la fel ca şi predictorii;
- în cazul în care controlul acestor factori nu permite să se obţină o măsură-criteriu stabilă, este posibil să fie întărită stabilitatea alegând un număr mai mare de măsuri-criteriu sau extrăgând eşantioane în mai multe ocazii.

A treia caracteristică a unei măsuri-criteriu este absenţa contaminării. Măsura-criteriu trebuie să fie efectuată în condiţiile în care factorii nepertinenţi nu pot influenţa scorul. Contaminarea este mai susceptibilă de a acţiona atunci când măsura-criteriu se prezintă sub forma unei scale de evaluare, care este în mod particular vulnerabilă la efectul de halo sau la tendinţa spre indulgenţă. Instrucţiuni foarte precise, ca şi o descriere detaliată a caracteristicilor de evaluat permit diminuarea riscului de contaminare. R.L. Thorndike a insistat asupra unui factor major de contaminare. Este vorba despre cazul în care scorul-criteriu al unui respondent este influenţat de cunoştinţele pe care le are evaluatorul cu privire la scorul predictor al acestui respondent, efect numit *predicţie în sine* (*self-fulfilling prophecy*). El semnalează că aceste două tipuri de contaminare produc aproape întotdeauna o corelaţie puternică între predictor şi criteriu.

Astfel, în cazul în care cineva vrea să ştie dacă un test de lectură în engleză permite prezicerea scorului la un test de fizică, evaluatorul, care atribuie note, cunoaşte scorurile studenţilor la testul de lectură (variabila-predictor) şi judecata sa poate fi influenţată de acest fapt. Riscurile de contaminare sunt probabil mai ridicate când un student a obţinut un scor ridicat la testul de lectură şi când el pare să prezinte o atitudine negativă faţă de fizică. Iată un caz în care evaluatorul poate fi puternic tentat de a cota favorabil un student, judecând că o atitudine aparent negativă a acestuia nu este decât o faţadă, decât fructul imaginaţiei sale sau pur şi simplu că acest student trebuie în mod normal să adopte o atitudine constructivă. Problema poate să apară şi atunci când un student obţine un scor slab la un test de lectură, demonstrând o atitudine favorabilă faţă de fizică. Astfel, evaluatorul poate să judece că acest student trebuie să fie penalizat printr-un scor slab. Soluţia problemei este foarte simplă: este necesar ca evaluatorul să nu cunoască scorurile respondenţilor la variabila predictivă.

O măsură-criteriu bună este aceea care prezintă avantaje practice, este simplu de utilizat, disponibilă şi necostisitoare. Pertinenţa, fidelitatea şi absenţa contaminării sunt caracteristicile esenţiale în alegerea unei măsuri-criteriu.

Se pot delimita diverse categorii de măsuri-criteriu :

- o primă categorie cuprinde măsurarea directă a producţiei sau randamentului, a volumului vânzărilor, a masei salariilor plătite, a numărului de clienţi pe oră sau pe zi, a numărului accidentelor pe lună sau pe an, a numărului de erori pe unitate, a mediei cumulate, a nivelului de şcolarizare etc. ;
- a doua categorie este constituită de randamentul la un test, precum examenele şcolare şi testele administrate de serviciul de personal al unei întreprinderi. Din această categorie fac parte scalele de evaluare (aceste prime două categorii nu sunt mutual exclusive ; de exemplu, media cumulată poate să aparţină uneia sau alteia dintre categorii, după modul în care se obţine) ;
- în a treia categorie, măsura-criteriu o reprezintă apartenenţa la un grup. Testele de personalitate, în particular testele de interes, sunt adesea considerate ca fiind valide, dacă permit o bună clasificare a indivizilor în grupe foarte precise. Testele de aptitudini sunt destinate să permită distingerea indivizilor care vor reuşi de cei care vor eşua într-o activitate sau într-un program de studiu dat ;
- a patra categorie o constituie măsurile-criteriu de natură statistică, precum coeficienţii de saturaţie, coeficienţii de funcţie discriminantă sau coeficienţii de consistenţă internă.

După enumerarea celor patru categorii de criterii, rezultă că un test nu poate fi mai bun decât criteriile utilizate pentru a stabili validitatea sa, deoarece toate slăbiciunile criteriului se vor reflecta în test, limitându-i totodată utilitatea. Dificultatea nu rezidă deci în alegerea unei măsuri-criteriu pertinente, ci în faptul de a nu alege decât una singură. Pentru ca un proces de validare să fie valabil, este necesar, de obicei, să se recurgă la mai multe măsuri-criteriu.

6.2.3.3. Coeficientul de validitate şi interpretarea sa

Coeficientul de validitate este, în esenţă, un coeficient de corelaţie. Cea mai utilizată metodă pentru a determina validitatea predictivă constă în calcularea corelaţiilor între scorurile la testul-predictor şi la măsura-criteriu. În procedura de calcul, distingem cinci etape :

1. alegerea unui eşantion de respondenţi ;
2. administrarea testului-predictor la respondenţi ;
3. aplicarea tratamentul matematico-statistic adecvat ;
4. colectarea datelor-criteriu ;
5. calcularea corelaţiile între valorile scorurilor-predictor şi valorile criteriului.

Coeficientul obţinut este unul de validitate. Este simbolizat prin r_{XY}, r fiind coeficientul de corelaţie, X predictorul şi Y criteriul. Orice factor care influenţează coeficientul de corelaţie poate să influenţeze coeficientul de validitate, iar datele trebuie să respecte postulatele specifice efectuării corelaţiilor. În cazul coeficientului Pearson, postulatul de liniaritate între cele două variabile este primordial. Cele două variabile trebuie să fie legate într-un mod liniar, în caz contrar valoarea coeficientului fiind subestimată. Chiar dacă în prealabil majoritatea variabilelor sunt legate liniar, este necesar să se verifice această legătură înainte de a calcula coeficientul. Pentru a face verificarea, este suficient

să se traseze graficul produs de două variabile într-un plan cartezian, fiecare axă reprezentând una dintre variabile. Dacă relaţia nu pare a fi liniară, scorurile se pot transforma într-o altă scală sau se poate utiliza o altă metodă corelaţională decât cea definită de K. Pearson, pentru a estima coeficientul de validitate.

În al doilea rând, trebuie luată în considerare întinderea (amplitudinea) diferenţelor individuale. Se ştie deja că o mică variabilitate a scorurilor limitează valorile pe care poate să le ia un coeficient de corelaţie. Atunci când se studiază validitatea, aceasta este o problemă majoră, deoarece se pot utiliza mai multe grupe de respondenţi. În fapt, este preferabil să se utilizeze un grup eterogen, adică un grup care prezintă o mai mare întindere a diferenţelor individuale, pentru a estima validitatea. Oferim, ca exemplu, ansamblul candidaţilor care solicită o funcţie dată. Când există un tratament după trecerea testului şi când grupul ales l-a epuizat, situaţia este mai riscantă. Dacă există o oarecare relaţie între scorul la test şi criteriu, grupul va produce o întindere restrânsă a variabilităţii prezise, coeficientul de validitate fiind şi el restrâns.

Există mai multe moduri de a interpreta un coeficient de validitate. Cel mai răspândit constă în a compara diverşi coeficienţi rezultaţi şi a alege testul care dovedeşte coeficientul cel mai ridicat pentru o situaţie dată. Un alt mod de a interpreta coeficientul de validitate este de a-l transforma într-un procentaj de varianţă. Astfel, ridicând coeficientul de validitate la pătrat, se obţine o proporţie de varianţă explicată prin acesta. De exemplu, la un coeficient de corelaţie 0,60, se va spune că 36% din varianţă sunt distribuite la cele două variabile sau că 36% din varianţa criteriului se datorează varianţei scorurilor la testul-predictor. Coeficientul de validitate trebuie să fie de aproximativ 0,7071 (adică rădăcina pătrată din 0,5) pentru ca jumătate din varianţa criteriului să fie explicată prin varianţa scorurilor la testul-predictor.

Amintim faptul că un coeficient de validitate mai mare de 0,70 este excepţional şi că el merită o examinare atentă atunci când este raportat. Într-o lucrare apărută în anul 1984 (*Survey Item Blank*), M. Smith propune o gradare a semnificaţiilor unui indice de validitate :

a) peste 0,50 – excelent ;
b) între 0,40 şi 0,49 – bun ;
c) între 0,30 şi 0,39 – acceptabil ;
d) sub 0,30 – slab (*apud* Armstrong, 1999, p. 399).

Un alt mod de a interpreta un coeficient de validitate ţine cont de caracteristicile grupului care va fi ales dacă testul va fi utilizat în timpul unui proces de selecţie. H.E. Brogden demonstrează că, plecând de la anumite postulate, coeficientul r_{XY} poate fi interpretat direct ca fiind o măsură a eficacităţii predictive, adică un raport între media scorurilor obţinute la o măsură-criteriu a grupului selecţionat şi aceea a altui grup. Dacă cineva are 200 de candidaţi care prezintă o cerere de admitere pentru un program de studii unde nu se acceptă decât 75 de studenţi, desigur că sunt preferaţi cei mai buni 75 de studenţi. Procedura optimă constă în a accepta toţi candidaţii pentru un anumit timp, pentru a proceda apoi la o colectare a datelor-criteriu, iar în final să nu se păstreze decât 75 de candidaţi, care prezintă scorurile-criteriu cele mai ridicate. Media scorului-criteriu a acestui grup „select" va fi, evident, cea mai ridicată printre toate mediile pe care cineva poate să le calculeze plecând de la fiecare combinaţie de 75 de candidaţi din 200. Această procedură este bună în teorie, dar, în practică, ea este costisitoare, chiar imposibil de aplicat.

Este mai rezonabil să se administreze testul de selecţie şi să se aleagă acei 75 de candidaţi care au obţinut cele mai mari scoruri la test. Coeficientul de validitate, cum

a sugerat H.E. Brogden, este egal cu raportul dintre randamentul mediu la criteriu al grupului de 75 de studenţi, admişi plecând de la testul-predictor, şi randamentul mediu obţinut de 75 de studenţi care ar fi fost aleşi dacă selecţia se efectua pornind de la scorurile-criteriu. Dacă, de exemplu, se obţine un raport $r_{XY} = 4/5$ înseamnă că grupul ales plecând de la testul-predictor va avea un scor-criteriu mediu egal cu 4/5 din acela al grupului determinat plecând de la scorurile-criteriu. Determinarea coeficientului de validitate prin această metodă, oarecum interesantă, prezintă un mare dezavantaj, deoarece presupune ca toţi candidaţii să obţină un scor-criteriu, ceea ce în multe situaţii este practic imposibil. Un raport r_{XY} egal cu 1.00 semnifică identitatea celor două grupe.

În fine, un alt mod de a interpreta coeficientul de validitate foloseşte eroarea de predicţie. Această interpretare presupune utilizarea ecuaţiilor de regresie.

Selecţia şi clasamentul. În acest caz, scopul este de a evalua şi interpreta validitatea unui test într-un context de selecţie şi de clasare. Un scor Y prezis, cu ajutorul unei ecuaţii de regresie, este scorul-criteriu mediu al respondenţilor care vor prezenta un acelaşi scor-predictor X. Persoanele vor avea rezultatele la criterii mai mari sau mai mici decât această medie prezisă. Eroarea de predicţie este diferenţa între scorurile-criteriu adevărate şi scorurile-criteriu prezise plecând de la ecuaţia de regresie. Variabilitatea acestor erori este măsurată de abaterea standard a distribuţiei erorilor. Trebuie să menţionăm că o serie de relaţii pot fi descrise prin intermediul unei ecuaţii liniare, altele nu. Un exemplu bine cunoscut este cel referitor la relaţia dintre anxietate şi randamentul şcolar, care este o relaţie neliniară. Cercetări experimentale au arătat că elevii prea mult sau prea puţin anxioşi au tendinţa de a avea un slab randament şcolar, iar cei moderat anxioşi prezintă un randament mai bun. Legătura dintre cele două variabile nu se poate descrie printr-o relaţie liniară simplă.

Avantajele abordării corelaţionale aplicate la validitate sunt numeroase. Metoda permite obţinerea unui indice cantitativ (sau coeficientul de validitate), care descrie relaţia dintre predictor şi un criteriu pentru un grup dat. De asemenea, permite să se prezică un scor-criteriu potenţial pentru fiecare respondent cu ajutorul ecuaţiei de regresie. Coeficientul de validitate poate fi interpretat în diverse moduri, dar aprecierea corelaţională reprezintă, cert, modul cel mai răspândit de a trata validitatea.

Această abordare prezintă şi inconveniente. Când relaţia dintre un predictor şi criteriu nu este liniară, trebuie utilizate tehnici speciale, câteodată complexe. O altă slăbiciune a acestei metode provine din faptul că un coeficient de validitate nu permite să se obţină o măsură directă a justeţei deciziilor, când luarea deciziilor este fondată pe utilizarea unui test. Este important ca această slăbiciune a coeficientului de validitate să fie luată în considerare, deoarece utilitatea unui test depinde de respondenţii selecţionaţi şi de proporţia respondenţilor care vor reuşi.

O modalitate adecvată de a evalua calitatea unui test, în cazul nostru validitatea, este de a aprecia justeţea deciziilor luate pe baza informaţiilor pe care le-a furnizat, mai exact proporţia deciziilor luate anterior şi care s-au dovedit a fi juste. Există o serie de indici care permit evaluarea validităţii unui test, plecând de la care vor trebui luate decizii de selecţie şi de clasare. Indicii sunt prezentaţi în continuare.

Indicele de eficacitate este o primă formă în care se poate exprima indicele de validitate. Este determinat de raportul dintre deciziile corect luate având ca bază rezultatele la un test şi numărul total de cazuri. Decizia corect luată este apreciată prin

cunoaşterea rezultatelor la măsura-criteriu obţinute de subiecţii cărora li s-a administrat testul. Cu cât proporţia deciziilor corecte este mai mare, cu atât testul este mai eficace sau mai valid, deoarece indicele de eficacitate poate fi considerat, de asemenea, un indice de validitate. Deciziile sugerate prin testul-predictor se împart în cel puţin două categorii care se exclud reciproc; în mod similar se procedează şi pentru scorurile obţinute la criteriu. Rezultatele de la test şi de la măsura-criteriu se transpun prin intermediul scorurilor (standardizate sau nu) într-o diagramă de corelaţie. Un exemplu este prezentat în diagrama următoare, conţinând date pentru 150 de subiecţi.

Tabelul 9

Diagrama de corelaţie între rezultatele la test şi cele de la criteriu

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
56						1			1					1			1			
55																1		1	2	
54										1							3	1		
53												1			1	2	4			
52														1		1				1
51														1	2	1				
50							1				1		2	4	1		1			
49								1		2		3	2	1	1					
48		1		1		3	1		2	2	4	3	1							1
47							1		3	3	6						1			
46		1				1		1	3	5							1			
45						1	1	4	5		1	1	1							
44							3	3	2			1								
43					2	1	5				1									
42						3				2										
41					3				1											
40				1		2				1					1					
39			2	1						1										
38		1		1	1				1	1										
37	1						1													

Liniile corespund măsurii-criteriu (valorile-criteriu sunt pe prima coloană, Y), coloanele corespund valorilor-predictor (valorile-predictor sunt pe prima linie, X). Prin considerarea valorilor de separaţie între subiecţii recomandaţi şi cei nerecomandaţi, atât pentru predictor, cât şi pentru criteriu, diagrama se poate structura în zone, după cum este reprezentată în tabelul următor.

Tabelul 10

Diagrama rezultatelor la test şi la criteriu, cu evidenţierea unei linii de separaţie

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
56						1			1					1			1			
55																1		1	2	
54										1							3	1		
53												1			1	2	4			
52														1		1				1
51														1	2	1				
50							1				1		2	4	1		1			
49								1		2		3	2	1	1					
48		1		1		3	1		2	2	4	3	1							1
47							1		3	3	6						1			

	46	45	44	43	42	41	40	39	38	37
24										1
25	1								1	
26								2		
27							1	1	1	
28				2		3				
29	1	1		1	3		2			
30		1	3	5						1
31	1	4	5		1	1	1			
32	3	3	2			1				
33	5				1					
34				2						
35			1							
36				1			1			
37				1						
38				1						
39				1						
40				1						
41										
42										
43										

Corelaţia Pearson între rezultatele prezentate este de 0,71. Rezultatele aparţinând celor patru zone ale tabelului astfel constituit se folosesc în calcule. În mod sintetic, datele pot fi expuse într-un tabel compus din patru câmpuri, identic cu cel necesar calculării coeficientului φ.

Tabelul 11

Tabelul celor patru câmpuri

		Teste-predictor		
		Eşec (–)	Succes (+)	Total
Rezultate la criteriu	Acceptat (+)	A (+)(–)	B (+)(+)	A+ B
	Refuzat (–)	C (–)(–)	D (–)(+)	C + D
Total		A + C	B + D	A+B+C+D

Patru celule prezintă o importanţă specială. Celula A conţine numărul de respondenţi acceptaţi după rezultatele măsurii-criteriu şi respinşi după administrarea testului-predictor (*pattern* + –). Ne mai putem exprima prin „recomandaţi", deoarece în majoritatea cazurilor de acest gen ne aflăm în situaţii de selecţie. „Recomandat" poate însemna certificarea prezenţei unui anumit nivel al unei aptitudini, caracteristici, trăsături (nivel considerat cel puţin satisfăcător). Exigenţa acestei „recomandări" creşte în funcţie de specificul examinării psihologice. Celula A formează grupul aşa-zişilor *falşi negativi*. Recomandarea criteriului este considerată ca atestând valoarea, purtând girul realităţii practice, cu condiţia să fie bine ales. Considerarea unui anumit criteriu înseamnă compararea cu „realitatea" a valorilor-predictor. Celula B conţine acel număr al respondenţilor care au fost recomandaţi de testul-predictor şi care au fost recomandaţi şi de măsura-criteriu (*pattern* + +). Concordanţa recomandării candidaţilor din partea celor două sisteme de evaluare face ca acest grup să fie numit al *pozitivilor*. Celula C cuprinde numărul acelor respondenţi care nu au fost recomandaţi de testul-predictor şi nici de măsura-criteriu (*pattern* – –). Concordanţa în privinţa recomandării ca subiecţi lipsiţi de valoare în privinţa unei anumite carateristici face ca acest grup să fie numit *al negativilor*. Celula D conţine numărul de subiecţi care au fost recomandaţi prin testul-predictor şi care au obţinut rezultate slabe la măsura-criteriu, ceea ce face ca grupul să fie numit al *falşilor pozitivi* (*pattern* – +). Celulele A şi D sunt neconcordante în privinţa recomandărilor. Practic, avem de-a face cu două feluri de erori, care pot fi comise la examinarea prin intermediul testului psihologic (şi la orice fel de examen): eroarea care constă în nerecomandarea celor ce posedă o valoare adevărată (celula A) şi eroarea care constă în recomandarea celor lipsiţi de valoare (celula D). Celulele B şi C au concordanţă în privinţa semnelor, conţin deciziile considerate exacte, pe când celelalte cupluri de rezultate sunt considerate inexacte. Un prim indice de eficacitate, denumit *indice al deciziilor corecte*, rezultă din raportul dintre suma celulelor conţinând numărul subiecţilor cu rezultate concordante şi numărul total de cazuri[1]:

$$P_{DC} = \frac{B+C}{A+B+C+D} = \frac{B+C}{N},$$

în care P_{DC} este indicele deciziilor corecte şi N numărul total de respondenţi (suma $A+B+C+D$). Indicele P_{DC} poate fi tratat ca un indice de validitate (ca grad de exactitate a bunelor decizii). P_{DC} este un indice foarte simplu de calculat, luând în considerare toate deciziile luate. Tipurile de decizii au aceeaşi pondere. Valoarea celulelor tabelului depinde de valoarea-limită a testului şi valoarea-limită a măsurii-criteriu. Luând ca bază datele prezentate în diagrama de corelaţie anterioară, rezultă următorul tabel:

Tabelul 12

Completarea celor patru câmpuri

		Teste-predictor		
		Eşec (–)	Succes (+)	Total
Rezultate la criteriu	Acceptat (+)	A = 5	B = 76	81
	Refuzat (–)	C = 22	D = 47	69
Total		27	123	150

1. O analiză aprofundată a deciziei în cazul examenului psihologic găsim la Horia Pitariu, *Psihologia selecţiei şi formării profesionale*, Editura Dacia, Cluj-Napoca, 1983.

În acest caz,

$$P_{DC} = \frac{B+C}{A+B+C+D} = \frac{B+C}{N} = \frac{76+22}{150} = 0{,}65 .$$

Deci indicele de eficacitate, în calitate de indice al deciziilor corecte, este egal cu 0,65. Indicele de eficacitate prezentat constituie proporţia bunelor decizii şi poate fi considerat un indice de validitate al testului.

Este uzuală şi folosirea unui alt indice, *indicele celor corect prognozaţi pozitiv*, pe care putem să-l notăm cu P_{CP} şi care se calculează prin formula:

$$P_{CP} = \frac{B}{B+D} = \frac{76}{123} = 0{,}62 .$$

Acest indice este mai pertinent când scopul procedurii de selecţie este de a maximiza numărul de persoane alese care vor reuşi. El reprezintă raportul între cei corect recomandaţi de testul-predictor (recomandaţi atât de test, cât şi la măsura-criteriu) şi totalul celor recomandaţi de testul-predictor.

Este interesant, de asemenea, să se calculeze *raportul de selecţie*, care este proporţia indivizilor recomandaţi prin testul-predictor din numărul total de candidaţi. În cazul nostru:

$$R_S = \frac{B+D}{A+B+C+D} = \frac{123}{150} = 0{,}82 .$$

Să presupunem că s-ar fi emis o exigenţă mai mare în privinţa valorii subiecţilor pe care ar fi trebuit să-i selectăm în condiţiile când valoarea de demarcaţie a criteriului ar fi rămas constantă. Linia de departajare a recomandaţilor pentru testul-predictor se deplasează, în acest caz, spre dreapta. Cu o asemenea dificultate ne confruntăm atunci când se cere pentru selecţie un număr relativ mic de candidaţi, existând un număr divers de situaţii din practica examinării care să justifice acest lucru.

Diagrama restructurată se prezintă în felul următor:

Tabelul 13

Rezultatele din tabelul 10, cu deplasarea liniei de separaţie

	24	25	26	27	28	29	30	31	32	33
56						1			1	
55										
54										1
53										
52										
51										
50							1			
49								1		2
48		1		1		3	1		2	2
47							1		3	3

	34	35	36	37	38	39	40	41	42	43
56				1			1			
55						1		1	2	
54							3	1		
53		1			1	2	4			
52				1		1				1
51				1	2	1				
50	1		2	4	1		1			
49		3	2	1	1					
48	4	3	1							1
47	6						1			

46		1				1		1	3	5
45						1	1	4	5	
44							3	3	2	
43					2	1	5			
42						3				2
41					3				1	
40				1		2				1
39			2	1						1
38		1		1	1				1	1
37	1						1			

							1			
1	1	1								
	1									
1										
				1						

Sintetizarea noilor date în tabelul celor patru câmpuri dă următorul rezultat:

Tabelul 14

Sinteza în tabelul celor patru câmpuri a datelor din tabelul 13

		Rezultate teste-predictor		
		Eşec (–)	Succes (+)	Total
Rezultate la criteriu	Acceptat (+)	A = 24	B = 57	81
	Refuzat (–)	C = 62	D = 7	69
Total		86	64	150

Indicele deciziilor corecte este: $P_{DC} = \frac{57+62}{150} = \frac{119}{150} = 0{,}79 .$

Se observă, aşadar, o creştere firească atunci când ridicăm pragul de recomandare la testul-predictor. Pentru al doilea indice, al celor corect prognozaţi pozitiv, vom avea următoarea valoare:

$$P_{CP} = \frac{B}{B+D} = \frac{57}{64} = 0{,}89 .$$

Observăm şi în acest caz o creştere, firească, a valorii indicelui, în condiţiile existenţei unei coeficient de validitate apreciabil. În condiţiile modificării expuse a valorii de recomandare a testului-predictor, vom avea următorul raport de selecţie:

$$R_S = \frac{B+D}{A+B+C+D} = \frac{64}{150} = 0{,}43 .$$

Observăm că ascensiunea indicilor precedenţi se face cu preţul scăderii raportului de selecţie, adică a procentului de subiecţi care vor fi recomandaţi de către testul-predictor din totalul celor participanţi la examinare.

Separarea grupelor. În cazul în care, după măsura-criteriu, se pot lua în considerare grupe separate de subiecţi (cel puţin două), compararea grupelor se poate realiza în mod uzual folosind un test statistic de semnificaţie a diferenţei mediilor scorurilor-predictor.

Concluzia testului arată dacă scorul-predictor mediu permite distingerea grupelor. Pentru exemplul anterior, se pot considera grupele de acceptaţi şi refuzaţi (determinate de valorile-criteriu). Să presupunem astfel că împărţim în două eşantionul de 150 de subiecţi: un grup cu valoarea-criteriu mai mare sau egală cu 46 şi un grup cu valoarea-criteriu mai mică decât 46.

Pentru semnificaţia diferenţei dintre cele două scoruri-predictor medii se poate utiliza testul τ. Statistica testului τ este:

$$\tau = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}},$$

unde prin N_2 s-a notat media grupului 1, adică media valorilor testului-predictor care au valori corespondente mai mari sau egale cu 46 la măsura-criteriu, cu X_2 s-a notat media grupului 2, adică media valorilor testului-predictor care au valori corespondente mai mici decât 46 la măsura-criteriu, cu s_2^1 varianţa primului grup, cu s_2^2 varianţa celui de-al doilea grup, cu N_1 numărul de subiecţi din grupul 1, cu N_2 numărul de subiecţi din grupul 2. Oferim, pentru exemplificare, următoarele date:

$$\overline{X}_1 = 35{,}59,\ \overline{X}_2 = 30{,}44,\ s_1^2 = 15{,}92,\ s_2^2 = 7{,}34,\ N_1 = 97,\ N_2 = 57.$$

Ipoteza nulă presupune că nu există diferenţe semnificative între mediile celor două grupe de subiecţi. Aplicând formula, obţinem următorul rezultat:

$$\tau = \frac{35{,}59 - 30{,}44}{\sqrt{\frac{15{,}92}{97} + \frac{7{,}34}{57}}} = \frac{5{,}15}{\sqrt{0{,}16 + 0{,}13}} = \frac{5{,}15}{\sqrt{0{,}29}} = 9{,}71.$$

Semnificaţia raportului τ poate fi identificată în tabela de distribuţie Student (t).

Din considerente statistice, semnificaţia unei diferenţe între grupe depinde de mărimea acestora, ceea ce, aparent, poate produce confuzie în anumite situaţii. Cu cât o grupă este mai mare, cu atât mai mult o diferenţă, chiar minimă, riscă să devină semnificativă. La limită, în cazul unor grupuri foarte mari, chiar dacă un test discriminează foarte puţin, este posibil ca o diferenţă infimă să fie declarată semnificativă statistic, iar testul să fie declarat valid.

Utilizarea unei alte metode de evidenţiere a validităţii unui test vizează studiul utilităţii sale, care presupune o analiză a costurilor şi avantajelor în diverse situaţii. Astfel, *indicele de utilitate* a unui test este determinat plecând de la avantajele care le reţin şi costurile care le generează.

Analize multiple. În această secţiune, sunt tratate pe scurt situaţiile în care se iau în considerare mai multe criterii şi/sau mai mulţi predictori[1].

Un singur predictor – mai multe criterii. Un singur test poate să utilizeze pentru prognoză un număr de criterii diverse. De exemplu, în activitatea de selecţie din diferite tipuri de organizaţii se foloseşte uneori un singur test, cel mai uzitat fiind un test de inteligenţă generală, ca fundament pentru selecţia de candidaţi. Aşadar, pe baza unui

1. Pentru o documentare completă, cititorul trebuie să consulte lucrări specializate în această problematică, de exemplu H. Gulliksen, *Theory of Mental Tests*, John Wiley & Sons, New York, Londra, Sydney, 1967.

singur test, sunt făcute prognoze pentru un număr de criterii diverse. În SUA, s-a practicat frecvent aplicarea de teste psihologice de admitere la colegiu, în scopul de a prognoza succesul viitor în respectiva instituţie de învăţământ. Desigur, nu se prognozează strict succesul în colegiu, ci succesul în diferite specialităţi. Nu avem o validitate, ci mai multe validităţi, în funcţie de criteriile multiple. Un singur predictor nu poate să dea informaţii unitare pentru diverse criterii. Din aplicarea formulei ecuaţiei de regresie rezultă clar că estimarea criteriilor este realizată, în ultimă instanţă, pe baza unei funcţii liniare din rezultatelor observate. De aceea, nu putem furniza informaţii diferite, răspunzând la întrebarea: *„Care activitate (dintre cele pentru care se face selecţia) este cea mai indicată pentru un candidat ?"*.

Mai mulţi predictori – mai multe criterii. Este o situaţie frecventă din practica testului psihologic, iar problema care se pune acum vizează necesitatea gestionării competente a unei baze de date mult mai voluminoase. Un exemplu tipic îl constituie testele de consiliere, în general, şi de consiliere pentru alegerea unei anumite instituţii de învăţământ pentru subiecţi, în special. Pe baza acestor teste, care sunt administrate la admiterea într-o instituţie de învăţământ, s-ar putea să avem la dispoziţie valorile de estimare diferenţială a succesului verosimil al studentului. Variabilele-predictor pot fi compuse dintr-un număr de diverse teste psihologice, care au fost administrate înainte de intrarea în respectiva instituţie de învăţământ, şi, pe lângă acestea, de o serie de note în diverse specialităţi de învăţământ. Din rezultatele la aceşti predictori, pe de o parte, şi din notele obţinute la un număr de diverse cursuri de specialitate (valori-criteriu), pe de altă parte, se poate calcula pentru subiecţi o matrice a ponderilor de regresie, modelele de regresie estimate având drept variabile dependente valorile-criteriu. Devine astfel posibilă precizarea, cu metodele indicate, a unui număr limitat de predictori, dintre cei consideraţi, care produc prognoze bune pentru un număr de criterii. O astfel de colecţie de predictori este foarte economică, deoarece poate fi utilizată repetat.

Un alt program din această categorie a fost dezvoltat în timpul celui de-al doilea război mondial. La Luftwaffe a fost compusă şi validată o baterie de teste, care a putut să prognozeze succesul la formarea de specialişti, de exemplu piloţi, navigatori etc.

În general, în sistemele sociale de specialitate pedagogică, militară sau industrială sunt stabilite un număr de diverse activităţi-criteriu. O analiză a acestor activităţi poate să indice faptul că numărul variabilelor de comportament sau al proceselor psihice, apreciate ca fundamentale sau primare, este limitat şi că măsura în care aceste variabile primare se asociază la diverse activităţi-criteriu variază, adică semnificaţia relativă a variabilelor se transformă de la o activitate la alta. Cum multe variabile sunt însă asociate, este posibil să se dezvolte un număr limitat de teste, care să cuprindă cele mai multe dimensiuni importante ale variabilelor de comportament necesare pentru succesul în diverse profesiuni. Avantajele sunt consilierea diferenţiată eficientă şi posibilitatea de a realiza clasificări.

Eficienţa unui set de predictori poate să fie mai mare, dacă adoptăm procedeul corelaţiei multiple. Pe lângă avantaje, un asemenea set prezintă şi o serie de limite. La mai mulţi predictori, corelaţia valorilor criteriilor observate şi estimate este mai înaltă decât pentru un singur predictor. Cu cât luăm în considerare mai multe variabile sau dimensiuni în setul de predictori, cu atât mai mare este verosimilitatea de a cuprinde multiple dimensiuni ale unui anumit criteriu special.

În cazul corelaţiei multiple, cu cât sunt introduşi mai mulţi predictori pentru un singur criteriu la un anumit eşantion, cu atât mai redus devine aportul datorat adăugării

unui nou predictor. Una dintre explicaţiile pentru această stare de fapt ar fi dificultatea de a dezvolta sau de a găsi teste suplimentare care să cuprindă componente ale criteriului neacoperite de testele deja existente.

Coeficientul de regresie. Coeficienţii de regresie sunt calculaţi în aşa fel încât variabilele combinate cu ajutorul lor din predictori, la un anumit eşantion, corelează maximal cu criteriul. Mărimea coeficientului de regresie depinde de importanţa relativă a unui predictor oarecare pentru estimarea criteriului. Toate încercările de a interpreta astfel coeficientul de regresie conduc însă la eroare. Este posibil să se dezvolte modele pe baza cărora să se poată estima contribuţia unei anumite variabile pentru criteriu, dar sunt necesare concepte speciale din matematica superioară. Când se încearcă interpretarea coeficientului de regresie, problema constă în aceea că testele-predictor la eşantioane singulare nu sunt, de regulă, independente unul de altul. Aceasta înseamnă că matricea de corelaţii între predictori, aflată la baza calculului coeficienţilor de regresie, nu conţine peste tot corelaţii nule. Când în eşantion toate variabilele corelează nul una cu alta, se poate arăta uşor că aportul unei anumite variabile la prognoza criteriului este egal cu pătratul corelaţiei respectivei variabile cu criteriul. Dacă ponderea regresiei este egală cu 0, ştim că variabila corespunzătoare nu contribuie cu nimic la prognoza criteriului. O pondere negativă (coeficient de regresie negativ) nu înseamnă că variabila corespunzătoare nu joacă nici un rol la prognoza criteriului. O variabilă-predictor poate să aibă o pondere de regresie negativă, în ciuda corelaţiei pozitive cu respectivul criteriu (în cazul regresiei multiple). Când o variabilă-predictor corelează scăzut cu criteriul, însă înalt cu unul sau mai mulţi predictori, este posibil ca această variabilă să aibă o pondere de regresie negativă. O astfel de variabilă poate să prezinte importanţă pentru prognoză, la fel ca o variabilă cu pondere pozitivă.

O astfel de situaţie s-a întâlnit în activitatea de selecţie a piloţilor americani în al doilea război mondial. Testele pentru aptitudini mecanice, de calcul şi de reprezentare spaţială corelau înalt pozitiv cu criteriul. Testul verbal avea o corelaţie foarte scăzută cu criteriul, însă prezenta o corelaţie destul de înaltă cu celelalte trei variabile. La calcularea regresiei multiple, a rezultat că ponderea de regresie pentru testul verbal era negativă. Corelaţia multiplă cu criteriul era înaltă când testul verbal era admis în bateria de teste decât în alte cazuri, cu toate că valoarea corelaţiei acestui test cu criteriul era aproape nulă. Fenomenul este uşor de explicat. Aptitudinile verbale de înalt ordin nu erau esenţiale pentru modul în care se făcea formarea în cursul celui de-al doilea război mondial, în armata americană. Ele erau necesare însă în testele creion-hârtie care trebuiau să diagnosticheze aptitudini mecanice, de calcul şi de reprezentare spaţială. Aptitudinile verbale erau necesare, în scopul înţelegerii instrucţiunilor şi itemilor în cele trei teste. Punctajul verbal prevăzut cu o pondere negativă serveşte la reţinerea sau la eliminarea aptitudinilor irelevante şi la scăderea lor din punctajul total.

Testele cu pondere negativă de regresie sunt numite *teste supresoare.*

Punctaje critice multiple (*cut-offs*). Această procedură nu este de recomandat pentru mai mulţi predictori, însă merită să fie amintită. Prin metoda *punctajului critic multiplu*,

care prezintă interes pentru prognoză în primul rând, se evidenţiază punctajele la un număr de diverse variabile-predictor. Un individ figurează drept recomandat pentru o anumită valoare-criteriu, dacă punctajul său la fiecare variabilă-criteriu depăşeşte o anumită valoare prestabilită. Acceptăm faptul că variabilele-predictor sunt măsurate în unităţi standard. Un examinat poate să fie recomandat pentru formarea ca pilot (criteriul) dacă rezultatele sale la un test de mecanică, un test de calcul sau un test de gândire logică sunt situate cel puţin la o unitate sigma deasupra valorii medii a unei grupe de referinţă. Într-un test de vocabular, din contră, este necesar să existe un singur punctaj, superior unei valori critice, care stă la o unitate sigma sub valoarea mediei. La un număr de diverse criterii, pot fi fixate diverse serii de valori critice pentru fiecare variabilă-predictor. La această metodă, se acceptă aprecierea că subiecţii care au un punctaj sub valoarea critică la un predictor, pentru o anumită formare sau o anumită activitate, nu sunt indicaţi. Ca atare, subiecţii al căror randament este superior valorii critice vor avea succes.

Procedeul nu s-a impus în practica de zi cu zi, din motivele expuse în continuare.

1. *Lipsa obiectivităţii în precizarea valorilor critice.* În prezent nu există nici o metodă statistică general acceptată pentru precizarea valorii critice optimale la predictori, în relaţie cu anumite criterii. Şi aceasta, probabil, din cauza dificultăţii de a determina analitic respectivele valori. S-ar putea ca, în mod natural, la fiecare variabilă-predictor să se stabilească prea exigent un punct critic, ceea ce face ca proporţia subiecţilor care depăşesc valoarea-predictor să fie foarte mică, deci cota subiecţilor respinşi la criteriu să fie prea mare. Şi aici trebuie definită însă *cota-parte a respinşilor.* S-ar putea stabili un punct critic pe baza părerii experţilor despre o cotă-parte de respingere sau de acceptare. Chiar când aceasta ar fi posibil, nu există siguranţa că o combinaţie de valori critice pentru mai multe criterii ar garanta o cotă de succes optimal. În această metodă, nu se acceptă faptul că o superioritate la o variabilă-predictor compensează o inferioritate într-o altă variabilă-predictor. Trebuie amintit că utilizarea ecuaţiilor de regresie presupune posibilitatea de compensare. Nu este o acceptare iraţională, deoarece din observaţii practice ştim că succesul depinde de diverşi factori şi, atunci când lipseşte ceva unui factor, aceasta se poate compensa printr-un alt factor. Analiza vicariantă se ocupă în special de asemenea fenomene de compensare.

2. *Lipsa unui indice corespunzător pentru efectivitatea acestui procedeu.* La regresia multiplă şi la alte procedee corelative se dă, o dată cu coeficienţii de corelaţie multiplă, un indice pentru efectivitatea prognozei, expresie a gradului de corelare a valorilor-criteriu observate şi estimate. În metoda punctelor critice multiple, acest lucru nu este posibil, deoarece candidaţii pentru un criteriu special, care au depăşit diverse punctaje critice, nu obţin nici o valoare diferenţiată şi nu se poate stabili cum diferenţiază efectiv acest sistem candidaţii aleşi în raport cu succesul.

Prognoza pentru mai multe criterii. Modelul de prognoză ideal conţine cât mai mulţi predictori şi mai multe criterii, el părând a fi cel mai economic şi mai cuprinzător pentru marile programe în care sunt folosite testele psihologice. Când există mai multe criterii, pot să fie aplicate două modele diferite. Unul oferă o prognoză multiplă diferenţială, celălalt o prognoză multiplă absolută.

Prognoza diferenţială multiplă are sarcina de a prognoza pe baza unei serii de predictori fiecare criteriu individual dintr-un număr de diverse criterii. Se poate decide atunci,

după estimarea valorilor-criteriu pentru fiecare subiect al eşantionului, activitatea-criteriu pentru care un candidat este cel mai recomandat. Se pune problema, în acest caz, de a alege predictorii care diferenţiază maximal indivizii cu privire la valorile-criteriu estimate.

Pentru rezolvarea acestei probleme, se acceptă că se oferă un număr relativ mare de teste-predictor. Se poate decide atunci ce teste trebuie alese, pentru a rezulta o diferenţiere cât mai bună între estimările valorilor-criterii. Aceste valori formează baza pentru recomandarea fiecărui candidat la o activitate, pentru care el este cel mai indicat, sau pentru o consiliere corespunzătoare.

Prognoza multiplă absolută a mai multor criterii se deosebeşte puţin de modelul discutat anterior. Ţelul prognozei multiple absolute nu este acela al diferenţierii maxime între valorile-criterii, cum este cazul la o prognoză diferenţială. Problema este de a le căuta, dintr-o mare cantitate de variabile-predictor, pe acelea la care suma corelaţiilor multiple cu un singur criteriu să fie maximă, fără luarea în considerare a corelaţiilor între criterii.

Ambele proceduri utilizează regresia multiplă, dar se deosebesc prin procedura de reţinere a unui test-predictor în vederea prognozei

Funcţia de discriminare multiplă. Funcţia de discriminare multiplă este un model aplicat la un set de criterii. Scopul metodei este acela de a discrimina maximal între un număr de diverse grupe de criterii pe baza unei mulţimi de variabile-predictor. De exemplu, într-o clinică psihiatrică pot fi prezente un număr de grupe de diverse tulburări psihice: schizofrenie, psihoză maniaco-depresivă etc. Pe baza unei astfel de diviziuni, trebuie găsite combinaţii liniare de predictori, discriminând subiecţii care nu sunt încă grupaţi. Această metodă poate fi utilizată şi dacă vrem să ordonăm muncitorii din diverse grupe de vârstă în vederea prognozării direcţiei de pregătire sau activitate.

Validitatea concurentă. Problema validităţii concurente sau concordante (în terminologia engleză, *concurent validity*) a unui test apare atunci când valorile-criteriu pot fi evidenţiate, practic, simultan cu punctajul testului. Comportamentul la test este valorizat aproape concomitent cu acel comportament exterior testului, fără ca distanţa temporală dintre ele să prezinte importanţă. Denumiri similare sunt: validitate simultană, concomitentă, convergentă, de statut. Tiffin şi McCormick (1967, p. 98) explică raţiunea ultimei titulaturi: „Validitatea de statut constă în a arăta în ce măsură rezultatele considerate la test sunt legate de un criteriu actual de muncă sau de un statut". Deşi similară, în spirit, cu validitatea de anticipare, procedura foloseşte în realitate postdicţia mai degrabă decât predicţia.

Validitatea concurentă poate fi stabilită atunci când la valorile unui test utilizat în psihologia clinică, ce îşi propune să diagnosticheze anumite tulburări de adaptare, sunt disponibile valori-criteriu rezultate din deciziile estimative ale psihiatrilor, profesorilor sau altor persoane competente în această direcţie. Scorul obţinut la un test de personalitate administrat de către un psiholog poate fi substituit rezultatelor la un examen dat cu o echipă de specialişti şi menit să detecteze prezenţa unei tulburări nevrotice, de exemplu. O astfel de substituire nu este valabilă decât dacă există o puternică relaţie între scorurile obţinute la test şi randamentul la criteriu şi dacă folosirea testului este rentabilă în materie de timp şi bani. Se vorbeşte despre validitate concomitentă atunci când rezultatele testului se pot substitui acelora ale evaluării randamentului la criteriu. Într-un asemenea caz, ne putem întreba din ce motiv mai trebuie aplicat testul, de vreme ce puteam avea

la dispoziţie o informaţie echivalentă, rezultată din practicarea unor metode clinice. Răspunsul este fără echivoc, în această privinţă: un test este mult mai economic decât alternativele clinice şi, în anumite situaţii psihodiagnostice, criteriul economic este foarte important, mai ales atunci când economia se realizează în timp. Coeficienţii de validitate obţinuţi în cadrul validităţii interne corespund mai degrabă aspectelor validităţii concordante. În acest ultim caz, scopul nu este de a prezice randamentul la criteriu, ci de a înlocui rezultatele evaluării randamentului la criteriu cu acelea ale testului.

O modalitate de a colecta valori-predictor şi valori-criteriu constă în formarea unei grupe experimentale de subiecţi pentru care sunt disponibile datele servind drept criterii. Există deci valori pentru aprecierea succesului în activitatea-criteriu. În general, avantajele validităţii concurente sunt dezavantajele validităţii predictive şi invers. Stabilirea validităţii concurente este preferabilă dacă pentru fiecare grupă de subiecţi implicaţi este disponibil un criteriu, după care se poate aplica imediat ecuaţia de regresie. Se pot aplica la această grupă o serie de teste-predictor, pentru ca, în scurt timp, să se dezvolte ecuaţii de regresie pentru estimarea criteriului din testele în cauză. Ecuaţiile pot fi aplicate imediat la noi angajaţi sau la cei admişi recent într-o instituţie de învăţământ.

Disponibilitatea imediată reprezintă un avantaj al validităţii concurente. Dezavantajele sunt mult mai numeroase. Din punct de vedere practic, este dificil să se aplice teste-predictor la subiecţi ce exersează activităţi ale căror rezultate cuantificate pot servi drept criteriu (timpul necesar pentru aplicarea testului se acordă mai greu de cei care conduc etc.). În al doilea rând, nu se poate şti cu exactitate când se hotărăsc subiecţii să participe la teste şi nu se ştie niciodată precis cât de în serios iau această participare. Elevii pot crede că rezultatele examenelor prin test pot fi folosite împotriva lor. Dacă nu le influenţează situaţia, s-ar putea ca ei să nu ia în serios efectuarea testului şi, ca atare, nu se recomandă folosirea rezultatelor, care pot fi considerate nereprezentative sau irelevante pentru aptitudinea ori procesul psihic de interes pentru cercetare. Există şi un dezavantaj teoretic. Nu putem fi siguri cât de mare este concordanţa între randamentul la test înainte şi după începerea activităţii-criteriu. Când există garanţia că între aceste două variabile corelaţia este relativ înaltă, atunci nu ar fi, din punct de vedere teoretic, nici un dezavantaj al valabilităţii concordante. Dacă însă efectuarea testului face apel la funcţii psihice aflate în strânsă relaţie cu criteriul, atunci influenţa exersării unor astfel de operaţii asupra rezultatului testului este destul de mare. Datorită acestui fapt, nu mai există o corespondenţă cu valorile aceluiaşi test, care s-ar fi obţinut dacă testul ar fi fost realizat în avans cu o perioadă de timp.

Validitatea de construct. Este definită ca indicaţie a gradului în care testul măsoară constructul măsurat. În bibliografia de specialitate, termenii similari utilizaţi pentru validitatea de construct sunt: validitate teoretică, validitate conceptuală, validitate ipotetico-deductivă.

L.J. Cronbach considera că fiecare test este, până la un anumit grad, „impur" şi măsoară foarte rar ceea ce se desemnează prin numele său. Testul nu poate fi interpretat atâta timp cât nu ştim prin ce factori este precizat punctajul său. Ideal ar fi ca, prin intermediul unei serii de precizări progresive, să se poată ajunge la o listă completă a surselor (izvoarelor, direcţiilor) care influenţează valoarea testului şi să se poată delimita intensitatea şi tipul fiecărei influenţe. Actualmente, nu se poate atinge decât relativ rar acest ideal. Interpretarea unui test se construieşte progresiv şi nu este niciodată deplină.

Considerăm sugestive o serie de observaţii de nuanţă pe care Grace H. Kent le face într-un studiu intitulat *Suggestions for the next revisions of the Binet-Simon Scale* (*apud* Cronbach, 1979, pp. 189-190), unde subliniază necesitatea de a preciza validitatea de construct.

> „Când un copil de vârsta cititului este trimis într-o clinică din cauza incapacităţii sale de a citi, este de primă importanţă să ne asigurăm dacă potenţialul său psihic se situează în limitele normale. O baterie de teste, care conţine material pentru citire, indică faptul că nu este vorba despre un subiect a cărui incapacitate de a citi se reduce la o altă cauză decât retardarea psihică. Un test care pretinde răspunsuri verbale delimitează foarte clar acest caz de acela al unui copil care, pe baza obstacolelor verbale, nu poate să se facă deplin înţeles. Este mai mult decât o farsă când se utilizează un test fără limită de timp sau un test cu itemi limitaţi temporal pentru subiecţi psihotici, ale căror procese psihice sunt încetinite patologic. Ceea ce vrem noi să măsurăm cu testul poate să fie foarte semnificativ, însă este cu totul altceva decât ceea ce trebuie să măsoare testul."

Se poate observa, în exemplul dat, că se încearcă diagnostigarea normalităţii psihice prin teste care verifică anumite capacităţi de citire sau de comunicare verbală, altfel spus, există o diferenţă marcantă între scopul declarat şi mijloacele utilizate.

Testul considerat inadecvat de Grace H. Kent corelează, probabil, cu criteriul succesului şcolar şi poate fi dovedit ca valid în privinţa conţinutului, deoarece prezintă un eşantion semnificativ de randamente de adaptare. Dificultatea constă în aceea că testul nu poate să fie interpretat ca măsură a unei singure calităţi psihologice.

Conceptul de *validitate de construct* a fost introdus oficial de American Psychological Association (APA) şi dezvoltat ulterior de către Comitetul pentru test al acestei organizaţii profesionale, avându-i ca reprezentanţi de frunte pe L.J. Cronbach şi P.E. Meehl (1955). Gustav Lienert consideră că dobândirea statutului de termen ştiinţific printr-o recunoaştere formală de către o societate naţională de psihologie îi conferă acestuia un caracter artificial. În opinia sa, integrarea termenului în teoria şi practica măsurării psihologice se va face foarte greu, dacă se lecturează cu atenţie dezbaterea pe marginea conceptului din publicaţia originală.

Termenul de *construct* a fost introdus în circuitul ştiinţific de fizicianul P.W. Bridgman, fondatorul operaţionalismului, curent idealist subiectiv al neopozitivismului. Cel mai simplu mod de folosire a termenului de *construct* este de a-l considera sinonim cu cel de *concept*, având în vedere faptul că ambele sunt creaţii logice, intelectuale[1]. Se foloseşte un *construct* când se încearcă stabilirea unei relaţii între mai multe obiecte sau evenimente. Când întâlnim sintagma „construct ipotetic", ne găsim frecvent în situaţia în care un domeniu de interes al cercetătorului nu este direct observabil sau măsurabil, însă este ipotetic, presupus că există, deoarece dă naştere unor fenomene măsurabile. *Ego*-ul este o astfel de construcţie teoretică. În concepţia lui L.J. Cronbach şi P.E. Meehl, noţiunea de *construct* este „suma atributelor ipotetice ale indivizilor, atribute care noi presupunem că sunt reflectate în rezultatele unui test" (*apud* Dorofte, 1981, p. 258).

1. Vezi pe această temă Arthur S. Reber, *Dictionary of Psychology*, Penguin Books, Londra, 1985, p. 151.

Jürgen Guthke dă următoarea definiţie constructului: „*Construct* este de cele mai multe ori înţeles în psihologie ca o caracteristică, respectiv complex de caracteristici, acceptată în primul rând ipotetic, care stă la baza comportamentului exterior observabil, el însuşi nefiind însă direct observabil" (Guthke, Böttcher şi Sprug, 1991, p. 291). Autorul definiţiei avertizează că prin utilizarea noţiunii de *construct* ne însuşim, cu sau fără voie, şi fundamentele idealist subiective ale acestui termen. Probabil autorul a rămas fidel unei distincţii făcute în filosofia marxistă (între filosofia idealist subiectivă şi filosofia idealist obiectivă, diferenţiere care pare, în prezent, revolută).

Considerăm necesare următoarele precizări:

- Constructul nu se confundă cu noţiunea sau conceptul (termeni sinonimi în logica tradiţională), dar pot exista segmente comune. El este mai degrabă un set de reprezentări asociate unui concept teoretic şi poate fi tratat ca o speţă intermediară între noţiunile teoretice şi noţiunile concrete sau empirice. Noţiunile teoretice nu pot fi corelate cu ceva ce există ca atare în realitate, ca lucru de sine stătător. Noţiunile concrete se aplică la lucruri care există (sau se presupune că există) ca atare în realitate. Noţiunile concrete redau o totalitate de determinări, iar cele abstracte – numai determinări separate sau reificate. Psihometricienii englezi utilizează termenul de *construct* pentru a desemna modelizarea unui concept. În acest sens, putem spune că testele nu măsoară INTELIGENŢA, ANXIETATEA sau INTROVERSIUNEA, ci ceea ce cred autorii testelor că sunt inteligenţa, anxietatea sau introversiunea.
- Constructul nu este direct observabil, nu-şi pierde calitea de abstractizare, însă există posibilitatea ca el să fie conectat cu fapte concrete, observabile. Un concept poate avea mai multe constructe. La rândul lor, constructele au mai multe *faţete*, având calitatea de componente independente. Analiza factorială este o metodă care poate ajuta la conturarea acestora. Elaborarea unui construct constituie un suport teoretic al testului psihologic.

Merită să precizăm că noţiunea de *concept psihodiagnostic* desemnează maniera generică de raportare a testului. Nici un test nu măsoară direct o trăsătură, ci într-un mod mediat, prin intermediul unui model al conceptului. Neglijarea acestei precizări a dus la reificarea unor concepte în faza de început a testării psihologice, cu consecinţe nefaste pentru aceasta. Astfel, inteligenţa era considerată ca având o realitate aproape de substanţă, care se localiza în creierul uman. Maurice Reuchlin (1992, pp. 51-52) spune în această privinţă: „Iluzia constă în a crede că, deoarece un cuvânt există în limbă, acestuia îi corespunde în mod necesar, fără ambiguitate, un obiect identificabil. Aceasta înseamnă să uităm pluralitatea definiţiilor posibile ale fiecărui cuvânt din limbă, faptul interpretării diverse a cuvintelor care compun definiţiile şi că limba nu are rolul de a furniza procedee empirice care ne-ar permite să ştim în ce măsură o anumită observaţie corespunde unei anumite definiţii (ea furnizează un cod, nu o metodă de codare)".

Validitatea de construct este tipul de validitate care pretinde un efort ştiinţific de un rafinament deosebit, deoarece se referă la natura instrumentului de măsură numit test psihologic. În psihologie şi în alte discipline, înţelegerea unui fenomen progresează în armonie cu capacitatea noastră de a-l măsura. Specialistul în măsurări se bazează pe o teorie cuprinzând trăsătura măsurată şi face predicţii privind comportamentul la test în mai multe situaţii. Aceste situaţii sunt verificate experimental şi, dacă ele se dovedesc exacte, se pronunţă validitatea teoretică a testului. Dar, dacă ele sunt false sau insuficient

de exacte, mai multe explicaţii se oferă acestui fapt: se poate ca experimentarea să fi fost un fiasco sau ca teoria de susţinere să fie falsă sau ca testul să nu măsoare constructul în chestiune. Realizarea măsurării presupune operaţionalizarea conceptuală, care stă în centrul atenţiei în cazul validităţii de construct şi care este un demers ştiinţific ce presupune o mare cantitate de conversiuni informaţionale; acestea produc, practic, o *traducere* a variabilelor teoretice în comportamente observabile. În cazul în care o ipoteză presupusă de construct nu se verifică şi instrumentul de măsură este bine operaţionalizat, teoria se cere schimbată sau ameliorată. Mircea Flonta (1976, p. 11) exprimă foarte clar această necesitate: „Eşecul unei teorii ştiinţifice în faţa testelor empirice, adică producerea unei contradicţii sistematice între consecinţele empirice deduse din teorie şi propoziţiile de observaţie, care exprimă rezultatele unor noi observaţii şi experimente, va atrage după sine abandonarea teoriei".

Validitatea de construct este considerată o clasare în cadrul unui sistem de referinţă teoretic al unui comportament manifestat (obiectivat) într-o situaţie de testare. Datele testului nu prezintă după aceste interpretări nimic altceva decât încercarea de operaţionalizare conceptuală a constructelor personalităţii. Respectivul proces al validării noţiunii este similar celui produs de descoperirea unei teorii sau a unui construct teoretic care face posibilă lămurirea unui comportament la test (H. Westmayer, 1972).

Noţiunea de validitate de construct presupune dezvoltarea unui eşafodaj teoretic, o reţea de concepte şi de rezultate experimentale. În evidenţierea validităţii de construct avem de-a face cu un dublu demers: unul teoretic şi unul empiric. Cel teoretic presupune parcurgerea bibliografiei de specialitate, cuprinzând cercetările experimentale anterioare în ceea ce priveşte precizarea aspectelor constructului. Accentul se pune şi pe rigurozitatea unor lanţuri de raţionamente ipotetico-deductive. Problema nevoii stringente de a schimba nivelul de exigenţă faţă de teoriile existente până acum câtva timp în domeniul psihologiei este abordată în literatura românească de specialitate psihologică de către Mircea Miclea (1994), care vorbeşte despre „presiunea de jos" venită din partea neuroştiinţelor în ultimele două decenii.

Semnalul de alarmă tras cu privire la calitatea teoriilor psihologice nu este o chestiune recentă, ci o preocupare durabilă a unor epistemologi de marcă. Karl Popper făcea următoarea remarcă, încă din anul 1934: „Pentru a putea clarifica această concepţie (deductivistă), este necesar să clarificăm opoziţia dintre psihologia empirică a cunoaşterii, care operează cu fapte empirice, şi logica cunoaşterii, care nu se interesează decât de corelaţii logice, căci prejudecăţile logice inductive sunt legate de confundarea problemelor psihologice cu cele epistemologice, ceea ce, în treacăt fie spus, are urmări neplăcute nu numai pentru teoria cunoaşterii, ci şi pentru psihologie" (vezi Popper, 1981, p. 75).

Dacă instrumentul de cercetare dovedeşte o operaţionalizare deficitară, atunci poate interveni schimbarea sau ameliorarea constructivă a instrumentului ori a procedurii de validare. Se ajunge adesea în punctul în care nu se pot măsura toate aspectele constructului, pentru că definiţia sa conţine elemente prea complexe sau prea vagi. Acesta este motivul pentru care validitatea teoretică se stabileşte mai curând printr-o acumulare de evidenţe decât printr-un singur indice.

Analiza psihologică a testului este chiar ţelul unei validări de construct, validare orientată mai mult teoretic decât pragmatic. Semnificaţia ei nu stă într-o valorificare practic-diagnostică nemijlocită, ci în clarificarea teoretică a ceea ce măsoară testul corespunzător. Cele mai multe dintre testele de uz curent nu sunt cunoscute după

punctele de vedere practic operaţionale, ci după caracteristicile acceptate, respectiv deschise, care sunt gândite ca aflându-se în spatele comportamentului diagnosticat în test (astfel, devine imediat clară semnificaţia validităţii de construct). Denumiri ca *teste de inteligenţă*, *teste de aptitudini* sau *teste pentru diagnosticarea altor caracteristici* tind clar la construct, deci la unităţi complexe, deduse, şi nu cuprinse nemijlocit operaţional. Din cauza acoperirii operaţionale deficitare, validarea de construct este un proces de o dificultate considerabilă şi cu mari costuri tehnice şi financiare. Teoria şi realitatea empirică se află, în acest caz, într-o strânsă relaţie reciprocă.

Testul şi constructul se află în relaţie reciprocă, prin aceea că testul poate să modifice constructul într-un anumit mod, iar constructul este asimilat în structura testului. În consecinţă, validitatea de construct se bazează într-o măsură considerabil mai mare pe cercetarea fundamentelor, în comparaţie cu validitatea logică şi cu cea orientată spre criteriu.

Validitatea de construct se rezumă la un set de proceduri de evaluare a validităţii unui instrument de testare, bazate pe determinarea gradului în care itemii testului cuprind trăsătura, constructul ipotetic pe care pretind că îl măsoară. Astfel, dacă se presupune că un test măsoară inteligenţa, ne putem întreba:

a) Ce trăsături sau calităţi caracterizează inteligenţa?
b) Itemii testului reflectă într-adevăr aceste constructe?

Validarea de construct cuprinde mai multe etape, posibil diferite de la un autor la altul, dar care în final realizează acelaşi obiectiv. Cronbach (1979, p. 191) ia în considerare trei etape:

1. Lansarea ipotezei în privinţa constructului care elucidează comportamentul la test. Acesta este un act de reprezentare, care se bazează pe observarea comportamentului la test şi pe cercetarea logică a testului.
2. Deducţia ipotezelor verificabile din teoriile localizate în construct. Aceasta este o operaţie pur logică.
3. Realizarea unei cercetări empirice pentru verificarea acestor ipoteze.

Jürgen Guthke (Guthke, Böttcher şi Sprung, 1991, p. 264) consideră că demersul pentru validare este constituit din patru trepte, pe care le redăm în rezumat:

1. Colectarea din fundamentele teoriilor asupra obiectului diagnosticului (de exemplu, anxietate, introversiune) a unei serii de aserţiuni, şi anume:
   a) despre relaţiile presupus pozitive ale constructului implicat în test şi alte constructe;
   b) despre neconcordanţele presupuse de relaţia dintre constructele cercetate şi alte constructe;
   c) despre relaţia dintre constructele cercetate şi anumite variabile observate (moduri de comportament, produse ale activităţii ş.a.m.d.).
2. Alegerea, mai ales pentru dezvoltarea unor teste, a acelora care apar adaptate pentru constructele indicate la „a" şi „b".
3. Formularea unei serii de ipoteze care se referă la măsurarea constructelor corespunzătoare (tipul „a" se referă la validarea convergentă, tipul „b" la cea discriminantă, iar tipul „c" la validarea criteriului).
4. Verificarea sau falsificarea, respectiv modificarea ipotezelor anterioare, pe baza cercetării.

Desigur, există şi alte clasificări, dar considerăm că autorii amintiţi conturează bine esenţa preocupărilor în privinţa etapelor de urmat pentru realizarea validităţii de construct.

Succesiunea faptică de operaţii nu este clar ordonată. Adesea se colecţionează informaţii cu un test, înainte de a oferi o interpretare.

Derularea cercetării poate cuprinde verificarea celorlalte validităţi. Într-o viziune extinsă a validităţii, cea de construct se relaţionează cu celelalte feluri de validitate amintite, uneori chiar înglobându-le. Pentru a stabili validitatea teoretică a unui test, toate felurile de predicţii sau întrebări pot fi formulate ori verificate empiric. Astfel, elementele următoare pot constitui criterii de conturare a validităţii teoretice: toate informaţiile precum natura itemilor testului, omogenitatea testului, stabilitatea sa în anumite condiţii, datele cuprinzând validitatea sa de conţinut şi validitatea predictivă, predicţiile relative la diverse grupe, modificările rezultatelor datorate timpului, corelaţiile între construct şi anumite variabile demografice şi, în final, procesul de măsurare utilizat pentru test.

Unii autori (D. Laveault şi J. Grégoire) consideră validitatea de construct ca fiind o validitate de conţinut extinsă, conferită unui test, de regulă, de către un consiliu de experţi, căruia i se cere să se pronunţe în ce măsură sarcinile sunt reprezentative pentru un concept ce constituie baza teoretică a testului. Itemilor li se studiază toate faţetele care converg spre identitatea noţională a anxietăţii, depresiei, introversiunii etc. Reproşul de subiectivitate poate fi contracarat prin crearea şi respectarea unei metodologii foarte riguroase. Validitatea de conţinut este direcţionată, în majoritatea cazurilor, spre conţinutul itemilor, care nu se confundă, de cele mai multe ori, cu conţinutul testului, pentru evaluarea căruia se iau în considerare şi alte aspecte, precum instrucţiunile date, modalităţile de răspuns şi de cotare, limita de timp.

Efortul de asigurare a validităţii de construct presupune intervenţia tuturor căilor metodologice posibile, în care se află în mod îndreptăţit una lângă alta abordarea experimentală şi cea empirică, analiza logică, analiza corelaţional-statistică. Verificarea practică a validităţii de construct necesită, frecvent, aplicarea metodelor statistice, dintre care metodele corelaţionale ocupă un loc important. Exploatarea acestor metode oferă o serie de coeficienţi sau indicatori ai validităţii testului. O bună parte dintre căile metodologice au fost prevăzute încă de la introducerea termenului de validitate de construct de către Cronbach şi Meehl, iar altele au fost introduse pe parcurs, cum s-a întâmplat cu analiza *multitrait-multimethod-Matrix*, iniţiată de D.T. Campbell şi D.W. Fiske, în care noţiunile de *validitate convergentă* şi *validitate discriminantă* joacă un rol decisiv.

Demersurile pot fi rezumate după cum urmează:

1. Corelarea testului cu un criteriu exterior.
2. Corelaţiile testelor cu testele care au exigenţe de validare asemănătoare.
3. Corelaţii cu testele care cuprind alte caracteristici ale personalităţii.
4. Analiza factorială a testelor de validat în comun cu criteriile exterioare, testele valid înrudite şi testele divergent valide.
5. Analiza diferenţelor interindividuale în rezultatele testelor.
6. Analiza transformărilor interindividuale la realizări repetate, cu şi fără variaţie sistematică a condiţiilor de realizare.
7. Analiza de conţinut logică a elementelor testului.
8. Studiul trăsăturilor latente.
9. *Multitrait-multimethod-Matrix*.

Toate cerinţele empirice amintite la punctele enumerate se completează reciproc, constructul aflat la baza testului de validat fiind analizat din diverse unghiuri. O mărime a nivelului de validitate de construct nu se poate preciza cu exactitate, consecinţă firească a utilizării unor metode diversificate.

Observăm că aproape toate punctele enumerate anterior au în comun utilizarea metodei corelaţiei. În acest caz este important şi necesar să ştim cu ce variabile ale altor teste este corelat constructul şi cu ce variabile nu este corelat. Un test menit să diagnosticheze posibilităţile de reprezentare spaţială are corelaţii pozitive cu testele spaţiale, conform naturii sale, şi nu cu cele verbale sau numerice. Este de aşteptat, cel puţin, ca valorile corelaţiilor cu variabilele „îndepărtate" de construct să fie considerabil mai scăzute decât cele existente cu variabilele „apropiate" de construct. D.T. Campbel şi D.W. Fiske propun, în 1959, ca la fiecare test să fie precizată o validitate convergentă (*convergent validity* – indicând corelaţia pozitivă a testului de validat cu alte teste) şi o validitate discriminantă (*discriminant validity* – indicând corelaţia nulă, lipsa corelaţiei cu alte teste).

Realizarea punctelor 1, 2, 3 şi întocmirea unui tabel cu astfel de rezultate cuprinzând ansamblul intercorelaţiilor fac relativ dificilă dobândirea unei perspective de ansamblu. În acest caz, analiza factorială poate să realizeze servicii adecvate. Rezultatele testelor, ale criteriilor îndepărtate de construct şi ale celor apropiate de construct sunt exploatabile prin intermediul analizei factoriale. Unul dintre factorii dobândiţi ar trebui să fie interpretat în sensul constructului. Testul de validat, testele valide utilizate şi criteriile ar indica în acest factor saturări înalte, iar cele îndepărtate de construct, din contră, încărcături scăzute. În mod normal, un test va indica saturare înaltă într-un singur factor, iar în ceilalţi ar trebui ca saturările să fie cât mai joase. Structura factorială a unui test este frecvent desemnată ca *validitate factorială* şi enumerată ca tip special de validitate. Din punct de vedere logic, ea este văzută ca prezentând numai un aspect al validităţii de construct.

Dintre exigenţele operaţiilor de la punctul 5 (analiza diferenţelor interindividuale) fac parte analiza distribuţiei rezultatelor (excesul, asimetria) şi verificarea unei diferenţe semnificative a valorii testelor între diverse grupe. Asemenea deosebiri între grupe pot să se refere la vârstă, statutul socioeconomic, formaţia şcolară şi alte criterii considerate relevante. Formarea acestor grupe se produce sistematic, corespunzător anumitor postulate deduse din construct.

În privinţa exigenţei formulate la punctul 6 (analiza transformărilor interindividuale la realizări repetate, cu şi fără variaţie sistematică a condiţiilor de realizare), se poate face observaţia că testul de validat este aplicat la aceiaşi subiecţi, în aceleaşi condiţii de realizare, astfel încât variaţiile constatate dau indicaţii mai ample despre validitatea sa de construct. Când un test se referă la o anumită caracteristică relativ constantă temporal, de exemplu inteligenţa, atunci el permite, conform propriei naturi, variaţii numai în interiorul limitelor sale. Pe de altă parte, un test pentru diagnosticarea situaţiei momentane a dispoziţiilor nu permite să fie temporal prea stabil, deoarece constructul *dispoziţie* implică o puternică variaţie relativă.

Repetarea testului nu trebuie, în mod necesar, să aibă loc în aceleaşi condiţii. Chiar sunt de recomandat repetări cu variaţii ale condiţiilor de cercetare, controlate experimental. Importante variaţii ale condiţiilor de cercetare sunt: transformarea instrucţiunilor, schimbarea examinatorului, modificarea nivelului de stres de la începutul probei, stări psihice diferite ale subiecţilor.

În privinţa exigenţelor de la punctul 7 (analiza de conţinut logică a elementelor testului), precizăm faptul că nu numai testul ca întreg poate să fie analizat din perspectiva constructului pe care se fundamentează, ci şi sarcinile, respectiv elementele testului, practic itemii testului, analiză care poate fi preponderent formală. Dintre aspectele formale fac parte, prioritar, intercorelaţiile şi analiza factorială a sarcinilor. Este raţional ca sarcinile să fie intercorelate pozitiv şi să fie atribuite, toate, unui singur construct. Consistenţa internă şi analiza factorială pot fi luate în considerare, în acest caz. În afară de aspectul formal, sarcinile testului pot fi analizate din punctul de vedere psihologic, adică se poate realiza o analiză de conţinut. Cu ajutorul autoobservaţiei şi al observaţiei, procesul de rezolvare la fiecare sarcină poate să fie făcut inteligibil şi să fie cercetat în compatibilitate cu constructul său.

Realizarea unei validări de construct ascunde anumite pericole. Deoarece nu se oferă o măsură clară a validării de construct, apare posibilitatea ca, pe baza corelaţiilor statistice, cercetarea empirică să fie respinsă în favoarea abordării speculative. Pentru diagnosticul psihologic pragmatic, argumentul corelaţiei între teste şi criterii este foarte util, aceeaşi atribuire putând fi făcută pentru lămurirea factorilor psihologici subiacenţi unui test. Pe baza cunoaşterii a ceea ce stă la baza unui test, se poate renunţa la aspectul pragmatic, după împrejurări, nu însă şi la validitatea referitoare la criterii. Corelaţia dovedită a unui test cu un criteriu bine definit îndreptăţeşte diagnosticianul care recomandă metodele şi procedeele sale ca utile şi raţionale.

Studiile de generalizabilitate sunt mai complexe, dar şi mai complete. Cea mai bună abordare este aceea a studiului validităţii *multitrait-multimethod* a lui Campbell şi Fiske. Reamintim că *validarea convergentă* permite să se demonstreze că două teste măsoară acelaşi construct, în timp ce *validarea discriminantă* constă în a demonstra independenţa sau noncorelaţia testului cu alte teste care măsoară un construct diferit. Prezentăm în continuare un exemplu de utilizare a tehnicii de analiză *multitrait-multimethod-Matrix* (*apud* Arias, 1995, p. 352).

Tabelul 15

Exemplificare pentru multitrait-multimethod-Matrix

		Metoda 1			Metoda 2		
		A1	B1	C1	A2	B2	C2
Metoda 1 Autoraportare	A1. Calitate lider	(0,95)					
	B1. Sociabilitate	0,28	(0,86)				
	C1. Popularitate	0,58	0,39	(0,92)			
Metoda 2 Aprecierea grupului	A2. Calitate lider	**0,76**	0,32	0,57	(0,95)		
	B2. Sociabilitate	0,30	**0,68**	0,40	0,39	(0,76)	
	C2. Popularitate	0,40	0,31	**0,60**	0,55	0,26	(0,74)

Sunt prezentate aici corelaţiile rezultate în urma aplicării mai multor metode, prin care s-au apreciat mai multe trăsături (în cazul nostru, două metode – autoaprecierea şi aprecierea de către grup – cu trăsăturile *calităţi de lider, sociabilitate şi popularitate).* Fiecare metodă constă din teste diferite, referitoare la aceeaşi caracteristică.

Prin această procedură nu este întotdeauna necesar să se compare acelaşi număr de constructe, de teste, de metode, fiind recomandate combinaţiile pertinente. Pentru

a putea fi exploatate, metodele au fost aplicate, firește, aceluiași grup de subiecți. În această matrice, coeficienții de fidelitate sunt trecuți între paranteze și cuprind coeficienții de corelație (r_{A1A1}, r_{B1B1}, r_{C1C1}, r_{A2A2}, r_{B2B2}, r_{C2C2}) realizați între aceleași teste din cadrul aceleiași metode. Pentru calculul lor, testele A1, B1, C1, măsurând trăsăturile A, B și C, sunt aplicate a doua oară. Se observă că sunt coeficienți destul de ridicați și stabili.

Coeficienții r_{A1B1}, r_{A1C1}, r_{B1C1}, r_{A2B2}, r_{A2C2}, r_{B2C2} sunt coeficienții de corelație rezultați din corelarea rezultatelor provenind de la aceeași metodă, dar aparținând unor teste diferite. R.M. Arias îi numește *coeficienți heterotrăsătură-monometodă.*

Cifrele îngroșate desemnează coeficienții de corelație între rezultatele la teste aparținând unor metode diferite, referitoare la aceeași trăsătură (r_{A1A2}, r_{B1B2}, r_{C1C2}). Sunt corelații statistice semnificative și suficient de ridicate.

Coeficienții de corelație r_{A2B1}, r_{A2C1}, r_{B2C1}, r_{A1B2}, r_{A1C2}, r_{B1C2} au rezultat din prelucrarea rezultatelor provenite din aplicarea unor teste aparținând unor metode diferite și referitoare la trăsături diferite. Aceștia sunt așa-numiții coeficienți *heterotrăsătură--heterometodă.*

6.2.3.4. Alte tipuri de validitate

Validitatea testului psihologic este raportată și printr-o serie de denumiri diferite decât cele prezentate până acum. Sunt validități de importanță mai mică, dar, în anumite contexte, pot să joace un rol important, în condițiile în care se asigură realizarea lor. Ele constituie, în mare parte, variante ale primelor trei sau proceduri simplificate determinate de contextul empiric. O parte dintre ele au fost amintite anterior, în rezumat sau cu un singur sens[1].

- Validitatea *a priori* (sau apriorică) este o estimare preliminară și intuitivă a validității de conținut a unui test. Ea stabilește gradul în care itemii testului au o legătură intuitivă *a priori* cu acele comportamente pe care se presupune că le testează. Se mai numește *validitate de simț comun.*

- Validitatea *congruentă*, amintită sumar anterior, este verificată printr-o metodă de stabilire a validității unui nou test prin corelarea scorurilor obținute la testul în cauză cu rezultatele altor teste, a căror validitate a fost stabilită anterior. Exemple pentru acest tip de validitate se găsesc în domeniul testelor de inteligență, unde noile teste create sunt comparate cu testele recunoscute ca fiind validate (cum sunt Stanford-Binet sau testele create de David Wechsler).

- Validitatea *diferențială* este validitatea stabilită pentru o baterie de teste și se referă la capacitatea acesteia de a prevedea diferențele de performanță a două sau mai multe criterii.

1. Aceste validități ocupă spații modeste în manualele sau lucrările de specialitate. Pentru inventarierea lor am consultat următoarele dicționare: Arthur Reber, *Dictionary of Psychology*, Penguin Books, Londra, 1985; *Grand Dictionnaire de la Psychologie*, Larousse, Paris, 1992; Wilhelm Hehlmann, *Wörterbuch der Psychologie*, Alfred Kroner Verlag, Stuttgart, 1968; Ursula Șchiopu (coord.), *Dicționar de psihologie*, Editura Babel, București, 1997; Helmuth Benesch, *Atlas de la Psychologie*, Librairie Générale Française, Paris, 1995; A. și R. Mucchielli, *Lexique de la Psychologie*, Entreprise Moderne d'Édition, Éditions Sociales Françaises, Paris, 1969; Henry Piéron (coord.), *Vocabulaire de la Psychologie*, Presses Universitaires de France, Paris, 1955.

Cota 239 pagina 13

• Validitatea *internă* este rezultanta unei proceduri interne de determinare a validităţii unui test, prin cercetarea fiecărei componente şi evaluarea gradului în care acestea îşi îndeplinesc rolul în cadrul testului. Astfel, se poate vorbi şi despre validitatea itemului, ca fiind gradul în care un singur item al unui test măsoară ceea ce a fost gândit să măsoare. Validitatea internă, numită uneori şi validitate *factorială*, ne indică în ce măsură itemii constituind testul sunt omogeni.

• Validitatea *incrementală* (*incremental validity*) este dată de informaţiile valide suplimentare, adiţionale, furnizate de un test, în afara celor obţinute prin alte procedee. A fost introdusă de Sechrest, în 1963, şi se aplică, în principal, într-un context de regresie multiplă în care sunt utilizaţi mai mulţi predictori. Un test (sau o variabilă) este valid atunci când el măreşte puterea predictivă a unui ansamblu de predictori. Acest tip de validitate pune accentul pe contribuţia particulară a unei variabile la puterea de predicţie.

• Validitatea *reprezentativă* (fondată pe un model) este o variantă a validităţii de conţinut, bazată pe evaluarea ponderii în care sunt reprezentate în test diferite caracteristici presupuse a susţine şi a fundamenta ceea ce se măsoară. Este mai mult decât o procedură de control, pentru a se evita subiectivismul în cadrul testului, şi are drept obiectiv prevenirea defavorizării adevăratei măsurări.

• Validitatea *sintetică* (*syntethic validity*) este validitatea unui instrument complex de testare sau a unei întregi baterii de teste şi se bazează pe relaţia dintre un scor compozit, care presupune reprezentarea diferiţilor factori prezenţi în test, şi performanţa reală. Termenul derivă din procedura reunirii mai multor factori într-o singură valoare. Procedura constă în a estima validitatea într-o situaţie precisă plecând de la o analiză logică a particularităţilor unei meserii sau ale unei specialităţi. Ea permite identificarea predictorilor potenţiali şi a criteriilor pertinente şi este folosită mai ales în aplicaţii din mediul industrial şi de afaceri.

• Validitatea caracteristicilor sau validitatea trăsăturilor (*trait validity*) este o componentă a unui proces mai vast de stabilire a validităţii de construct şi vizează determinarea gradului în care caracteristicile fundamentale, pe care ar trebui să le măsoare testul, sunt într-adevăr măsurate. Este apropiată de validitatea nomologică (vezi mai jos).

• Validitatea *intrinsecă* desemnează validitatea unui instrument de testare bazată pe itemii testului, indicând clar faptul că aceştia evidenţiază, într-adevăr, caracteristica diagnosticată. De exemplu, un item din test, care solicită celui examinat să silabisească un cuvânt, este un item intrinsec valid într-un test referitor la abilitatea de a silabisi.

• Validitatea *aparentă* sau validitatea de faţadă (*face validity*) este aceea evaluată de către profani (ei având, în această situaţie, calitatea de „experţi") care apreciază conţinutul unui test pentru a vedea dacă este adecvat trăsăturii psihice pe care pretinde că o măsoară. Validitatea aparentă se apreciază în mod subiectiv, în acest caz putându-se crea impresia că testul este practic util sau că este fundamentat din punct de vedere teoretic. Deci semnifică faptul că un instrument de măsură, sub forma aparativă sau creion-hârtie, apare profanilor ca un bun indicator pentru ceea ce încearcă să prognozeze. Dacă vrem să prognozăm succesul pentru formarea piloţilor, atunci testul trebuie să conţină mici imagini cu avioane (chiar la sarcinile pentru posibilităţile de reprezentare spaţială, s-ar putea utiliza avioane stilizate în diverse poziţii). Profanul, cu semnificaţia de persoană neiniţiată sau nespecializată în teoria şi practica testului, poate avea şi calitatea de subiect

de examinare prin intermediul testului. Validarea aparentă poate fi o procedură destul de vagă de validare a unui test şi datorită subiectivităţii inerente a celui pus să aprecieze, fiind tipică doar fazelor incipiente de construcţie a testelor. Trebuie să remarcăm că, în timp ce validitatea aparentă pare, superficial vorbind, similară cu validitatea de conţinut, procedurile sunt destul de diferite, ultima fiind o procedură sistematică, iar prima fiind mai apropiată de validitatea *a priori*. Acest tip de validitate adaugă o dimensiune neacoperită prin alte tipuri de validitate. Validitatea aparentă cuprinde direct conţinutul şi structura testului. Când se spune despre un test că posedă o validitate aparentă, se deduce că există o corespondenţă logică şi evidentă între itemi şi ceea ce un test este destinat să măsoare.

Acest fel de validitate a beneficiat de o atenţie specială de-a lungul timpului. În unele cazuri, este considerată un tip principal de validitate, alături de validitatea predictivă, validitatea concurentă, validitatea de conţinut şi validitatea de construct[1]. Validitatea aparentă este lipsită de relevanţă pentru validitatea reală, faptică, empirică. Nu este necesar ca un test psihologic să apară ca şi când ar cuprinde aceleaşi aptitudini care sunt necesare la activităţile-criteriu. Asemănarea pare să fie de folos când intră în relaţie cu publicul interesat. Multe cercetări asupra testelor psihologice implică necesitatea ca participantul, personalul însărcinat cu implementarea testului (din partea beneficiarilor competenţei psihologilor) şi a funcţiilor prevăzute să fie convinşi de validitatea testelor. De îndată ce acest fapt este acceptat, examinarea poate începe. La testele din domeniul militar, generalii şi amiralii trebuie să fie convinşi de validitatea testului, la testele din industrie trebuie să fie convinşi preşedinţii, vicepreşedinţii şi directorii, la testele de admitere în învăţământul superior trebuie să fie convinşi decanii, şefii de departamente. În aceste situaţii, poate fi de folos validitatea aparentă. În experienţa de psiholog, dobândită în domeniul industrial şi al transporturilor, am observat deseori că managerii unor firme doresc să vadă testele aplicate în unităţile pe care le conduc şi, la fel de frecvent, să le efectueze. O validitate aparentă deficitară poate să hotărască soarta unor contracte de examinare prin intermediul testelor. Să nu uităm că Robert Yerkes a putut să testeze milioane de oameni pentru că a convins şefii militari de utilitatea acestor teste. În domeniul marinei militare a SUA din acele timpuri nu s-au realizat testele, deoarece şefii militari nu au fost convinşi de utilitatea lor. Dacă validitatea aparentă nu este compensată prin celelalte tipuri de validităţi, există pericolul utilizării unor teste lipsite de valoare. Un test poate să aibă o corelaţie foarte scăzută cu un criteriu, cu toate că, aparent, conţine moduri de comportament relevante pentru criteriu.

Deşi acest tip de validitate prezintă un interes limitat, credibilitatea testelor în rândul publicului larg depinde în mare măsură de validitatea de faţadă sau validitatea aparentă.

• Validitatea *nomologică* este o variantă a validităţii de construct, care derivă din argumentul că validitatea de construct implică două eforturi separate : unul se concentrează asupra validităţii constructelor (adesea numită validitate a caracteristicilor) şi unul asupra validităţii testului într-un cadru teoretic mai larg, aceasta fiind cunoscută ca validitate nomologică.

• Validitatea *de consens* sau de părere generală se realizează printr-o procedură informală de evaluare a validităţii, bazată pe noţiunea de părere generală. Aceasta porneşte de

1. Vezi în acest sens Michael Armstrong, *A handbook of human resource management practice*, Kogan Page Ltd., Londra, 1999, p. 398.

la aserţiunea că o afirmaţie este cu atât mai posibil să fie adevărată cu cât sunt de acord mai multe persoane în privinţa adevărului ei – altfel spus, cu cât sunt de acord mai multe persoane în privinţa unei percepţii, cu atât este mai posibil ca ea să fie adevărată. Termenul de validitate de consens este rar folosit în cercetarea cu teste şi măsurători. Este un termen tipic pentru psihologia socială şi se conformează principiului conform căruia susţinerea sistematică a unei anumite poziţii îi acordă o mai mare acceptabilitate din partea celorlalţi.

• Validitatea *de eşantion* este o varietate a validităţii de conţinut, propunându-şi să evalueze gradul în care trăsăturile variate presupuse a se afla la baza a ceea ce este măsurat sunt reprezentate în test. Este mai mult o procedură de control, de prevenire a introducerii unei deformări sistematice (*bias*) în test, decât o măsură reală a validităţii testului.

Observăm că aceste variante de validitate au referinţe specifice şi prezintă considerabile interferenţe de sensuri.

6.2.3.5. Exigenţe faţă de validitatea unui test

În interpretarea coeficientului de fidelitate, se pot face afirmaţii categorice. Cu totul altfel stau lucrurile în legătură cu coeficienţii de validitate, care nu utilizează în nici un caz norme rigide, fiind indicate numai liniile directoare, în funcţie de care este apreciat un coeficient de validitate empiric. O apreciere relativă are mai mult realism şi mai multă valoare decât una absolută, pentru un domeniu structurat complex. Din punctul de vedere al prognozei statistice, trebuie să fie pretins un coeficient de validitate mai mare sau egal cu 0,7. Exprimarea cantitativă, prin intermediul coeficienţilor de validitate, are relevanţă mai ales în cazul validităţii referitoare la un criteriu. În cazul validităţii de construct, pot fi activate o serie de procedee cantitative şi calitative, ansamblu greu sintetizabil într-un indice sau coeficient.

Vom face câteva consideraţii asupra reperelor de apreciere a validităţii:

- Un test trebuie să fie valid într-o asemenea proporţie încât aplicarea să facă o mai bună afirmaţie decât neglijarea sa. Există mai multe cercetări care indică faptul că testele cu coeficientul de validitate în jur de 0,3 prezintă importanţă practică. În principiu, trebuie să ne mulţumim cu acei coeficienţi de validitate care permit să se asigure statistic un grad de încredere de 1%, indicând un adevărat raport între test şi criteriu. Testele cu un coeficient de validitate sub 0,3 sunt aproape fără utilitate în aplicarea individuală, chiar şi în situaţia în care sunt asigurate statistic. În cadrul unei baterii de teste putem realiza, din contră, un bun serviciu utilizând un astfel de test, mai ales când el corelează scăzut cu alte teste din baterie.
- Valoarea pretinsă unui coeficient de validitate depinde mult de domeniul de aplicaţie a testului psihologic. De exemplu, pentru activităţile de consiliere a persoanelor individuale (expertiza individuală) în care un psiholog trebuie să se pronunţe cu privire la gradul de reliefare a unei anumite caracteristici a personalităţii sau a unei aptitudini pentru ocuparea unui post, se pretinde un $r_{tc} > 0,70$ (coeficient de corelaţie test – criteriu), în contextul în care decizia în privinţa subiectului depinde exclusiv sau aproape exclusiv de rezultatul testului. În practica aplicării testelor, un coeficient de validitate de un asemenea nivel este obţinut în mod excepţional. Acesta este motivul pentru care se apelează la bateriile de teste, care sunt combinaţii adecvate de teste individuale pentru un anumit scop psihodiagnostic, cu şanse mult mai mari de a îndeplini exigenţele consilierii. Validitatea mai înaltă trebuie să fie cerută unui test atunci când de rezultatul obţinut de un anumit candidat depinde o decizie importantă, deci când o decizie greşită ar fi foarte dezavantajoasă pentru candidat.

- În cazul în care, pe lângă rezultatele testelor, sunt atrase şi alte informaţii necesare în evaluarea caracteristicilor individuale (şi acest lucru se întâmplă, de regulă), exigenţele faţă de coeficientul de validitate se reduc la valoarea de $r_{tc} > 0,5$ (în funcţie de ponderea acestor informaţii suplimentare). În consultarea individuală, interesul psihologului se orientează exclusiv spre subiectul respectiv.
- Un caz diferit îl constituie activitatea de selecţie profesională prin intermediul examenului de aptitudini. În acest caz, psihologul ia în considerare, în primul rând, interesul declarat al beneficiarilor, care pot fi organizaţii, firme, instituţii etc. Se ajunge în situaţia de a solicita selecţia a cât mai mulţi candidaţi prezumaţi ca fiind de succes, interesul centrându-se pe detectarea contraindicaţiilor clare. La întrebarea „Cât de înaltă trebuie să fie, în acest caz, validitatea pentru o astfel de selecţie?", răspunsul depinde de raportul stabilit între oferta de candidaţi şi nevoia de forţă de muncă a întreprinderii, ca şi de cota-parte procentuală a celor indicaţi în grupa candidaţilor. În nici un caz un coeficient de validitate nu poate scădea sub 0,30.
- În cazul comparaţiei între grupe, coeficientul de validitate, în raport cu coeficientul de fidelitate (care poate să fie scăzut în anumite cazuri de excepţie), trebuie să fie relativ înalt. Numai o astfel de valoare poate da siguranţa că s-a studiat deosebirea în privinţa caracteristicii cercetate şi că este vorba despre acea caracteristică pe care trebuia să o măsoare testul, şi nu de alta, pe care o măsoară, practic, în locul acesteia.
- Un test nou creat trebuie să posede o validitate mai înaltă decât un test utilizat până în acel moment, pentru acelaşi domeniu de validitate. Dacă există, de exemplu, un test de integrare şcolară, care are un coeficient de validitate de $r_{tc} = 0,5$, testul nou creat va trebui să depăşească această valoare, bineînţeles luând în considerare şi pragul de semnificaţie. Excepţie de la acest principiu se face doar atunci când noul test este considerabil mai economic decât cel vechi (de exemplu, timp mai scurt, mai puţin material folosit, costuri financiare mai reduse) şi are un coeficient de validitate egal cu testul utilizat anterior.
- Validitatea unui test trebuie să fie invers proporţională cu economia sa. Un test individual cu o durată de administrare de 90 de minute trebuie să fie mult mai valid decât un test de grup cu durata de 20 de minute.
- Unui coeficient de validitate îi este pretinsă, în general, o valoare mai înaltă atunci când coeficientul de fidelitate al testului se află la limita de jos a acceptării. În acest caz, testul are, prin extindere, şanse bune de a-şi îmbunătăţi valoarea.
- Exigenţele faţă de nivelul validităţii cresc o dată cu perioada temporală, pentru care trebuie să fie valabile afirmaţiile. La prognozele pe termen lung intervine, în mod natural, un factor de învăţare, care este variabil individual şi scade siguranţa prognozei.
- Pentru activitatea de prognoză sunt deosebit de importante validitatea de criteriu şi validitatea practică. În schimb, pentru diagnoza care presupune mai multă înţelegere

şi clarificare, sunt importante validitatea logică, validitatea factorială, descrierile pe care se bazează acestea fiind mai bogate în explicaţii decât validitatea practică. Este adevărat că validitatea logică exclude stabilirea unui coeficient de validitate.

- Studiile iniţiale ale construirii unui test sunt direcţionate, de obicei, spre validitatea de construct. Se ştie că acest tip de validitate este greu exprimabil exclusiv cantitativ. Spre deosebire de o validitate depinzând de măsurătorile de corelaţie, aici nu există fundamente matematice unice pentru determinarea validităţii. În comparaţie cu măsurătorile care depind de structurile de comportament cunoscute (de exemplu, validitatea legată de criterii), exprimarea validităţii de construct este întotdeauna în schimbare, o dată cu acumularea mai multor cunoştinţe despre trăsăturile şi calităţile ce fundamentează constructul.

6.3. Criterii secundare de apreciere a unui test psihologic

6.3.1. *Normativitatea unui test*

Acest criteriu presupune, ca exigenţă a unui test, posibilitatea de a oferi informaţii care să poată servi la ordonarea rezultatelor individuale ale testului în raport cu un sistem de referinţă, şi anume sistemul normativ. La testele unidimensionale, normarea decurge firesc din tendinţa de a afla distribuţia valorilor testului unei populaţii de subiecţi, de a transforma această distribuţie în una normală şi de a fixa anumiţi parametri ai distribuţiei (de exemplu, scala coeficienţilor de inteligenţă cu o medie de 100 şi o abatere standard de 15). Fiecare valoare brută a testului face parte dintr-un anumit nivel valoric al testului care fixează univoc poziţia unui subiect pe o scală.

Normarea poate să se realizeze pentru întreaga populaţie a unui areal (norme complete) sau pentru populaţia unui anumit grup social (norme de grup). Se pot fixa norme după grupe de vârstă, după sex, după proveniența urbană sau rurală, după nivelul de studii. De asemenea, se pot combina criteriile de separare a subiecţilor, în vederea realizării de norme, de exemplu: grupa de vârstă 25-35 ani, cu studii generale şi de proveniență urbană etc.

Un test poate îndeplini o serie de criterii, dar, dacă nu este normat, nu are nici o utilitate diagnostică, sau doar una foarte redusă, cu toate că este indicat ca instrument de cercetare (acolo unde este vorba, de regulă, numai despre comparaţii între grupe).

Procedeele de normare a rezultatelor obţinute la teste de către grupele de referinţă sunt tratate într-un capitol ulterior, referitor la aspectele constructive.

6.3.2. *Standardizarea unui test*

Într-o serie de lucrări consacrate teoriei şi construcţiei testului, standardizarea este prevăzută ca o cerinţă fundamentală, deoarece nerespectarea acestei exigenţe ar duce la serioase pierderi de informaţii sistematice. Eşantioanele de informaţii dobândite, de regulă, din comportamentul observat al examinatului variază în funcţie de caracteristicile psihice ale examinatorului şi de condiţiile necontrolate, care apar sistematic sau nesistematic.

La modul ideal, standardizarea are rolul de invariant al condiţiilor de examinare, în derularea procesului psihodiagnostic, şi face posibili indicatorii nedeformaţi ai caracteristicii

B.C.U. „M. EMINESCU" IAŞI

psihice în cauză. În practică, stabilirea influenţei sistematice asupra rezultatului testului se produce doar condiţionat.

Standardizarea procesului psihodiagnostic prin intermediul testului se face în mai multe faze: provocarea stimulării, înregistrarea răspunsului, evaluarea rezultatului şi interpretarea lui. Această fracţionare a procesului de diagnosticare prin test este o exigenţă teoretică, dar, în examinarea concretă, fazele nu pot fi separate la fel de eficient la toate clasele de teste.

În faza de provocare a stimulării, pe baza configuraţiei de itemi, se constată variaţii neînsemnate ale ordonării experimentale, care pot influenţa mai mult sau mai puţin rezultatele testului.

O serie de atenţionări pot juca un rol benefic, şi anume:

- păstrarea doar a instrucţiunilor de cercetare proiectate pentru test;
- utilizarea materialelor de cercetare recomandate pentru test (altfel, rezultatele nu pot fi comparate);
- crearea aceloraşi condiţii pentru toţi participanţii, deoarece, pe lângă materiale, relevante sunt şi indicaţiile date subiecţilor pentru situaţia de testare;
- necesitatea ca instrucţiunile şi condiţiile să nu aibă aceeaşi pondere, ci să fie diferenţiate;
- cerinţa ca un test să fie realizat la fel pentru toţi candidaţii, într-o ambianţă pe cât posibil neperturbată.

Condiţiile exterioare nu pot fi identificate complet, deoarece intră în calcul starea vremii, raportul spaţial, starea instrumentului de scris, eventualele mirosuri indezirabile, idiosincrasiile perceptive. Atenţia pentru crearea unei situaţii neperturbante, în privinţa condiţiilor de ambianţă, nu poate fi comparată cu cea din medicină.

Mai greu de delimitat şi normalizat sunt condiţiile intraindividuale, proprii fiecărui subiect. Obţinerea unor concluzii psihodiagnostice din rezultatele unui test presupune ca subiectul să se găsească, în momentul examenului, într-o dispoziţie corporală şi psihică normală.

Perturbările momentane ale stării de sănătate, efectele farmacologice pe durată scurtă sau lungă, tensiunile emoţionale acute, grijile, gradul de neliniște, nesiguranţa, oscilaţia dispoziţiilor personale etc. pot să influenţeze în mai mare sau mai mică măsură cantitatea şi calitatea răspunsurilor.

Atenţie specială trebuie acordată efectelor posibile care rezultă din dispoziţia personală, din aşteptările personale ale examinatului faţă de cercetarea prin test şi din consecinţele presupuse ale acesteia.

Cercetări empirice arată că anumiţi determinanţi speciali (motivaţia efectuării testului, participarea eului, anxietatea faţă de test, succesul sau insuccesul trăit, caracterul stresului – presiunea examenului şi a timpului) pot să influenţeze negativ calitatea diagnosticului.

În sens mai larg, strategiile de realizare, care variază interindividual, înclinaţiile latente sau ascunse la colaborare, minciuna, exagerarea şi afirmaţiile neconvingătoare, ca şi stilul de reacţie general sau specific al candidatului pot constitui surse de perturbări interne, psihice, provocând rezultate neregulate ale testului.

Provocarea comportamentală, ca situaţie socială, este prezentă mai ales la examenul individual, la care participă şi un complex semnificativ de condiţii marginale. Printre acestea se numără caracteristicile randamentului la test, vârsta, sexul, statutul social,

vestimentaţia, caracteristicile presupuse sau aşteptate ale examinatorului, mecanismele de interreacţie condiţionate (prezenţa sau absenţa examinatorului în spaţiul de examinare, cunoaşterea sau necunoaşterea lui, intensitatea observării şi a sprijinului, condiţionările atmosferice, mimice, gestuale sau verbale, încurajarea sau descurajarea, lauda sau mustrarea). Anna Anastasi spunea, în 1968, că semnificaţia variabilelor situaţionale la examenul prin test are valoare diferenţială în funcţie de fiecare test.

6.3.3. *Înregistrarea răspunsurilor*

Cu privire la înregistrarea răspunsurilor la test, se poate spune că aceasta nu este niciodată deplină. Există o diferenţă între înregistrarea rezultatelor şi observarea liberă a datelor. Înregistrarea cunoaşte forme multiple şi presupune protocolarea verbală a răspunsului, depozitarea anumitor semnale prin utilizarea de mijloace tehnice : aparat de fotografiat, casetofon, instalaţii video, aparate de înregistrare numerice şi verbale, terminale ale computerelor etc.

Înregistrarea standardizată a datelor trebuie realizată în măsura în care aceasta nu se produce automatizat. În acest caz se stabilesc reguli clare, în aşa fel încât fiecare examinator să înregistreze aceeaşi cantitate de informaţii.

6.3.4. *Evaluarea răspunsurilor la test*

Prelucrarea rezultatelor înregistrate la teste, precizarea frecvenţei anumitor reacţii--răspuns coerente se realizează parţial cu diverse aparate de evaluat, mai ales într-un test de aptitudini. Testele cu răspuns creativ fac imposibil lucrul cu maşinile de evaluat. Ţelul evaluării datelor este obţinerea unui rezumat al datelor înregistrate, având ca rezultat unul sau mai mulţi indicatori. Valorile brute se pretează la transformări pe baza raportării la datele grupurilor de referinţă. Evaluările sunt însoţite, în unele cazuri, de calcule asupra sindroamelor şi de prezentări grafice în profiluri, pentru a face datele mai uşor de înţeles. În evaluare, se fac simţite efectele de informare, precum şi de halo, de centrare, de contrast, de ancoraj etc.

6.3.5. *Interpretarea rezultatelor*

Calitatea consecinţelor diagnostice pentru interpretare constituie obiectul de interes al obiectivităţii. Dintr-un rezultat individual, mai mulţi cercetători trebuie să tragă aceleaşi concluzii. În timpul examenului însă, se produc o serie de interacţiuni între examinator şi examinat. Cu cât regulile sunt mai puţin formalizate, cu atât mai mare este riscul efectelor de interpretare explicată prin ecuaţia personală a examinatorului. Imperfecţiunile sunt doar incomplet corijate. Faţă de exigenţele unei valorificări depline există obiecţii mai mult sau mai puţin îndreptăţite, datorate caracterului de „realitate deficitară" al variaţiei condiţiilor controlate empiric.

Eşantioane nereprezentative de subiecţi şi metodele inadecvate de analiză a datelor îşi pun amprenta asupra calităţii interpretării. În practică, este posibilă doar o estimare a condiţiilor marginale, care anormalizează situaţiile de standardizare a fazelor individuale din procesul psihodiagnostic. Ele sunt stabilite segmentat chiar la testele construite şi verificate cu grijă.

Pentru asigurarea unei bune interpretări sunt necesare, uneori, aplicări ale unor teste preliminare suplimentare, care să limiteze efectele deformatoare (pentru stabilirea motivaţiei testului sau scalelor de minciuni).

În cazul rezolvărilor parţiale ale testelor, pot fi falsificate fundamentele de interpretare.

6.3.6. *Comparabilitatea unui test*

Un test este comparabil dacă :

a) există una sau mai multe forme de teste paralele ;
b) sunt disponibile teste asemănătoare în privinţa validităţii.

Forma paralelă a unui test permite comparaţia testului cu sine însuşi şi face posibil un control intraindividual al fidelităţii, prin care se cercetează un anumit subiect cu o anumită formă de test şi se compară rezultatele. Testele asemănătoare sau egal valide examinează aceeaşi caracteristică a personalităţii sau una apropiată.

6.3.7. *Economia unui test*

Un test este economic atunci când :

a) presupune un timp de realizare scurt ;
b) are nevoie de puţin material ;
c) se aplică simplu ;
d) este realizabil ca test de grup ;
e) este repede şi uşor de evaluat.

Un test este foarte economic dacă îndeplineşte toate sau cele mai importante dintre aceste condiţii. El este foarte puţin sau deloc economic dacă îndeplineşte doar o parte sau nici una dintre exigenţele enumerate. Nu a fost introdus, până în prezent, un indicator sintetic pentru aprecierea economiei unui test.

6.3.8. *Utilitatea unui test*

Un test este util dacă măsoară o caracteristică a personalităţii pentru a cărei cercetare ia naştere o necesitate practică. Un test are o mare utilitate dacă nu poate fi reprezentat în funcţia sa printr-un alt test. El are o utilitate redusă dacă examinează o caracteristică a personalităţii care ar putea să fie la fel de bine cercetată cu o serie de alte teste.

6.4. Relaţiile reciproce între criteriile unui test psihologic

Aceste relaţii reciproce s-au bucurat de un interes constant din partea psihometricienilor. De respectarea lor depinde posibilitatea utilizării practice a unui test. Gustav Lienert (1967, pp. 19-20) indică următoarele raporturi între criteriile unui test psihologic :

- Fidelitatea testului paralel sau retest poate să nu fie mai înaltă decât consistenţa lui sau decât obiectivitatea. Indicatorii de validitate pot să fie mai mici decât cei de fidelitate.

- Un test cu o înaltă validitate trebuie să posede în mod necesar şi o înaltă obiectivitate, consistenţă sau încredere. Stabilirea unei înalte validităţi ne scuteşte într-o anumită măsură de reverificarea criteriilor de apreciere cunoscute.
- Un test cu o validitate suficientă şi o consistenţă redusă trebuie să posede o încredere ridicată. Un asemenea test are mai multe şanse de îmbunătăţire, deoarece consistenţa sa poate fi sporită uşor prin anumite tehnici de construcţie a testului (în timp ce, invers, o lipsă de încredere îl face lipsit de valoare). Un test cu o fidelitate redusă se recomandă în cel mai bun caz pentru comparaţia grupelor în privinţa caracteristicii de cercetat. Pentru aceasta este suficientă o fidelitate de 0,5, nu însă şi pentru o diferenţiere individuală.
- Un test cu o slabă validitate şi cu o fidelitate înaltă trebuie să prezinte o încredere redusă. Fără o prelucrare de conţinut, nu există pentru acest test nici o posibilitate de îmbunătăţire. Se recomandă în diferenţierea indivizilor, însă nu oferă nici o certitudine pe baza căreia se produce această diferenţiere.
- Un test cu o validitate redusă şi o fidelitate îngustă este, practic, lipsit de valoare.
- Coeficientul de validitate nu poate depăşi rădăcina pătrată a coeficientului de precizie. Astfel, validitatea este întotdeauna limitată de fidelitatea unui test.

7. Analiza de itemi

7.1. Itemii ca elemente ale măsurării psihometrice

Dintre autorii unor lucrări de specialitate privind formele alternative ale itemilor, avantajele şi dezavantajele fiecărei forme, regulile care trebuie respectate în construcţie, precum şi ansamblul tehnicilor de prelucrare, îi enumerăm, selectiv, pe R.M. Travers (1950), J.K. Bennett şi J.E. Doppelt (1956), L. Guttman şi I.M. Schlesinger (1967), H.G. Osburn (1968), G. Lienert (1969), R.I. Thorndike şi E. Hagen (1977), J.S. Ahmann şi M.D. Glock (1971), D. Magnusson (1975), D.A. Payne (1974), R.K. Hambleton şi R. Jones (1993), D. Laveault şi J. Grégoire (1997)[1]. Regulile de exploatare a informaţiilor dobândite din domeniul itemilor se diferenţiază în funcţie de tipul de test folosit. În modelul clasic, sunt folosite valorile testului ca indicatori pentru reliefarea cantitativă a caracteristicii psihice. În majoritatea cazurilor, numărul şi succesiunea itemilor testelor psihometrice sunt fixate, deşi strategia secvenţială, dependentă de informaţie, în sensul tehnicilor *Item sampling* (preocupând un număr apreciabil de autori, dintre care amintim pe F.M. Lord şi M.R. Novick,1968, R. Linn, 1967, N. Burton şi K. Remer, 1972, W.E. Loadman, 1973, L.F. Hornke, 1976, 1977, 1978, G.H. Fischer şi P. Pendl, 1977, şi W. Buchtala, 1977), pune la îndoială o astfel de stabilire. Determinările, legate de item, ale valorilor testelor au la bază, în general, abstracţii care limitează considerabil spectrul mostrelor de itemi. Valoarea brută a testului se defineşte ca sumă a valorilor egal ponderate ale itemilor, ceea ce implică posibilitatea substituirii reciproce a itemilor omogeni.

D. Laveault şi J. Grégoire (1997, p. 223) compară analiza itemilor cu repetiţia unei orchestre în care instrumentele trebuie să cânte armonios şi să intervină într-un mod foarte precis: „Totul trebuie să producă o senzaţie muzicală aparte corespunzând

1. Indicăm între paranteze anii de apariţie a unor lucrări despre teoria itemilor.

intenţiilor compozitorului şi ale dirijorului". Mesajul furnizat de itemi – spun autorii citaţi – trebuie să fie clar, precis şi armonios.

Am amintit deja că procesul de măsurare cu ajutorul testelor psihometrice se bazează pe informaţii elementare provenite din domeniul itemilor. În primul rând, considerăm că interesează definiţia itemului şi precizarea faptului dacă toate testele sunt compuse din itemi.

O definiţie clară a itemului este dată de Maurice Reuchlin: „Element al unui test, constituind o situaţie particulară şi jucând rolul unui stimul la care subiectul nu poate da decât un răspuns dintre două sau mai multe răspunsuri teoretice prevăzute de test"[1]. Itemul este deci o componentă de stimulare a unui test care beneficiază de un dispozitiv de răspuns cu alegere multiplă şi care are menirea de a contribui la evidenţierea unei trăsături, a unei caracteristici psihice, pe care dorim să o diagnosticăm.

Când există o singură variantă de răspuns corect, dintr-un număr de *n* variante oferite, atunci celelalte *n–1* variante se numesc *distractori*. Pentru a-şi îndeplini rolul funcţional, într-un sistem de răspuns cu mai multe variante, distractorul trebuie să prezinte verosimilitate (aceasta constituie o direcţie distinctă de cercetare în domeniul psihodiagnosticului).

Numărul de itemi ai unui test este foarte variabil, dar limita inferioară nu scade sub 10, valoare considerată de Paul Kline (1994, p. 103) ca fiind limita pentru care un ansamblu de itemi se poate numi test. Fiecare item posedă un scor care, împreună cu scorurile celorlalţi itemi, permite exploatarea cantitativă a informaţiilor furnizate de test. Scorul este punctajul acordat răspunsului exprimat de către subiect. La probele de randament psihic, în special cognitiv, dacă răspunsul subiectului la solicitarea unui item este corect, scorul va fi egal cu 1, iar dacă acesta este incorect, scorul va fi egal cu 0. În cazul aplicării unui chestionar de personalitate, calitatea răspunsurilor nu se apreciază prin termenii de *corect* şi *incorect*. În acest caz, valoarea 1 o obţine un răspuns care pune în evidenţă caracteristica investigată prin respectivul chestionar, iar valoarea 0 – în situaţia contrară. Itemii sunt caracteristici mai ales pentru testele din domeniul cognitiv, din cel al personalităţii şi al cunoştinţelor, care se prezintă, în majoritatea cazurilor, sub forma *creion-hârtie*. Pentru a fi considerată item, o componentă a testului trebuie să aibă individualitate informaţională pertinentă, adică să contribuie distinct, clar conturat, la scorul total al testului. Rezultatele unui item pot fi prelucrate independent de rezultatul global al testului (autonomia informaţională a itemului). Bararea unei litere O sau C, la testul Bourdon-Amfimov, nu aduce o informaţie cu individualitate, deşi această operaţiune se punctează. Ca atare, un asemenea răspuns nu constituie un item. Nu poate fi vorba despre itemi nici la proba aparativă DM_1 a lui Lahy, folosită intens în laboratoarele psihologice din domeniul transporturilor pentru diagnosticarea asocierii şi disocierii manuale.

Noţiunea de *item* nu este unanim preluată de către autorii care se ocupă de teoria testului psihologic. De exemplu, David Magnusson, un autor care nu şi-a „americanizat" terminologia ştiinţifică, foloseşte (1975, pp. 22-35) noţiunile de *sarcini* şi *analiză de sarcini*, în loc de *itemi* şi *analiză de itemi*, deşi referirile se fac clar la ceea ce se numeşte analiză de itemi. Gustav Lienert (1967, p. 93) nu foloseşte, pe parcursul tratatului său privind construcţia testelor psihologice, noţiunea de *item* decât o singură dată, şi atunci pentru a indica o procedură similară de analiză a discriminării, întâlnită la A.F. Vukovich.

1. Vezi *Grand dictionnaire de la psychologie*, Larousse, Paris, 1992, p. 406.

Luată în sens general, analiza itemilor este o analiză detaliată a itemilor individuali dintr-un test sau un chestionar, cu scopul de a evalua validitatea şi siguranţa, gradul de încredere acordată fiecăruia dintre itemi, în general disponibilitatea testului pentru practică. Se poate vorbi despre *o analiză a itemului* şi o *analiză a itemilor*, cea dintâi trimiţând la aplicarea unor proceduri tehnice licite de prelucrare informaţională şi a unor proceduri semantice de modificare a conţinuturilor transmise, iar cea de-a doua la tehnicile de analiză vizând satisfacerea de către ansamblurile de itemi a unor exigenţe fixate anterior. Şi în acest domeniu, întregul reprezintă mai mult decât suma părţilor (exemple de probleme specifice ansamblurilor de itemi sunt cele referitoare la efectele de succesiune a itemilor).

Sub aspect general, asemenea analize pot fi axate pe conţinut şi formă sau pot fi realizate în mod cantitativ, observându-se modul în care fiecare item contribuie la precizia măsurării şi la validitatea generală a testului. În sens restrâns, analiza de itemi este folosită mai ales pentru evaluarea modului efectiv în care fiecare item contribuie la validitatea generală a testului.

Analiza nu poate fi făcută decât pentru testele la care subiecţii sunt lăsaţi să parcurgă toţi itemii, deci este imposibil de realizat la aşa-numitele *teste de rapiditate* (*speed tests*), adică teste cu timp limitat de administrare şi construite în aşa fel încât foarte puţini subiecţi pot termina toate solicitările în perioada de timp fixată. Este adevărat că o serie de autori susţin ideea potrivit căreia la testele realizate incomplet de către subiecţi se pot prevedea celelalte răspunsuri neefectuate prin studierea atentă a modului în care subiectul a răspuns până atunci. Considerăm că o asemenea previziune ar aduce un plus de relativism în privinţa concluziilor, relativism prezent oricum în interpretare. În cazul în care un mare număr de persoane nu au putut să răspundă la item din lipsă de timp, indicele de dificultate nu este exprimat cu adevărat.

Analiza de itemi este diversificată, opţiunea pentru o anumită categorie depinzând de scopul special urmărit în construcţia unui anumit test. Caracteristicile valorilor brute ale întregului test sau ale itemilor individuali nu sunt independente de caracteristicile eşantioanelor care sunt măsurate cu ajutorul testului. Acest aspect al eşantionării îl vom trata în capitolul referitor la construcţia testului. Un model teoretic ireproşabil, care să integreze perfect caracteristicile subiecţilor din eşantion şi caracteristicile variabilelor măsurate, nu este disponibil. Creatorii de teste construite pe baza teoriei probabiliste a testului psihologic aspiră la un model care să ofere parametri ai testului în mod independent de eşantioanele cercetate şi chiar de testele utilizate. În această lucrare, vom face, în mod deosebit, aprecieri referitoare la o metodă de analiză pentru un singur eşantion. Acest eşantion trebuie să fie reprezentativ şi pentru alte grupe la care testul urmează să fie aplicat. Abordarea pragmatică a metodei analizei de itemi începe prin trecerea în revistă a câtorva dintre scopurile sale cele mai importante. Se pot distinge două feluri de analize de itemi: acelea la care, în urma prelucrării rezultatelor, se evidenţiază un singur indice şi acelea la care sunt relevaţi mai mulţi indici pentru selectarea şi recomandarea unui anumit item.

Realizarea unor exigenţe abstracte, bazate pe o puternică formalizare cantitativă, este o operaţiune dificilă şi conduce, în cazul statisticilor neexhaustive, la o pierdere de informaţii despre posibilele configuraţii ale categoriilor de itemi. Pe de altă parte, formalizarea şi raportarea la modele constituie proceduri de neevitat, dar postulatele pot fi, în unele cazuri, „slăbite". Astfel, modelul teoriei clasice a testului oferă fundamentul pentru dobândirea valorilor fidele ale testului. Putem subscrie în final la ideea potrivit

căreia testele psihometrice se caracterizează prin aceea că numeroasele categorii posibile de itemi sunt ordonate şi delimitate fie pe calea avansării unor categorii selective, fie prin categorizări ulterioare, după prescripţii standardizate. La categoriile de itemi ordonate numeric, valorile numerice inferioare indică o reliefare scăzută, iar valorile superioare denotă o evidenţiere puternică a caracteristicii analizate. În majoritatea cazurilor, sunt preferate structurile informaţionale binare, cu categorii de măsurare 0 şi 1, dar acest model nu este obligatoriu de urmat, existând şi alte posibilităţi. Dacă o valoare a unui test trebuie să indice sub formă de măsură caracteristica pusă în discuţie, atunci aceasta presupune că şi categoriile itemului indică relaţii sistematice cu categoriile caracteristicii de măsurat.

Unul dintre scopurile analizei de itemi este de a selecta dintr-o colecţie mai mare de itemi (existentă în faza preliminară a construcţiei unui test) una sau mai multe grupe de itemi, în aşa fel încât punctajele corespunzătoare ale testelor să satisfacă diverse condiţii prestabilite cu privire la distribuţie şi la anumite corelaţii. Ca tipuri de analize de itemi putem enumera analiza dificultăţii, a distractorilor, a omogenităţii, a specificităţii, a validităţii, a excesului şi a asimetriei.

În construcţia testelor psihologice, se pleacă de la un ansamblu de itemi mai numeros decât cel care se consideră necesar, deoarece pe parcursul perioadei de elaborare a testelor unii itemi pot fi eliminaţi, ca urmare a neconcordanţei lor cu exigenţele unor criterii de valabilitate.

Structura de colectare a datelor necesare în constituirea scorului total al testului a fost creată de William Stern în anul 1912, purtând denumirea de *matricea lui Stern*. Exprimată în formă generică, ea se prezintă în felul următor:

Tabelul 16

Matricea Stern prezentată în formă generică

	Item						
Subiect	**1**	**2**	**...**	**i**	**...**	**N**	**t_j**
1	X_{11}	X_{12}	...	X_{1i}	...	X_{1n}	$\sum_{i=1}^{n} X_{1i}$
2	X_{21}	X_{22}	...	X_{2i}	...	X_{2n}	$\sum_{i=1}^{n} X_{2i}$
3	X_{31}	X_{32}	...	X_{3i}	...	X_{3n}	$\sum_{i=1}^{n} X_{3i}$
4	X_{41}	X_{42}	...	X_{4i}	...	X_{4n}	$\sum_{i=1}^{n} X_{4i}$
...	...	...	...	...	...	...	...
J	X_{j1}	X_{j2}	...	X_{ji}	...	X_{jn}	$\sum_{i=1}^{n} X_{ji}$

	Item						
Subiect	**1**	**2**	**...**	**i**	**...**	**N**	**t_j**
...	...	...	...	...	...	...	$\sum_{i=1}^{n} X_{Ni}$
N	X_{N1}	X_{N2}	...	X_{Ni}	...	X_{Nn}	$\sum_{i=1}^{n} X_{Ni}$
f_i	$\sum_{J=1}^{N} X_{j1}$	$\sum_{j=1}^{N} X_{j2}$	...	$\sum_{j=1}^{N} X_{ji}$	...	$\sum_{j=1}^{N} X_{jn}$	$\sum_{J=1}^{N}\sum_{i=1}^{n} X_{ji}$

Valorile din coloane cuprind rezultatele subiecţilor la un anumit item, cele din linii desemnând rezultatele unui subiect anume la totalitatea itemilor la care a răspuns. Analiza în coloană poartă numele de *ipsatizare*, iar cea în linie a valorilor – *obiectivare*. Cu f_i am notat suma valorilor pozitive ale itemului, adică a valorilor 1, care au semnificaţia rezolvării corecte.

Prezentăm în continuare cele mai frecvente tipuri de analize existente în acest domeniu.

7.2. Tipuri de analize şi indicatori

7.2.1. *Analiza dificultăţii*

Este cel mai simplu şi mai răspândit tip de analiză care se poate realiza la o serie de valori ale itemilor. Cu p_i se notează în mod obişnuit *indicele de dificultate a itemului*, indicându-ne proporţia subiecţilor care au rezolvat corect itemul. Formula sa este:

$$p_i = \frac{f_i}{N},$$

unde f_i reprezintă suma valorilor obţinute de toţi subiecţii la un anumit item, iar N numărul total de subiecţi care au răspuns la itemul respectiv. O valoare apropiată de 1 indică un item uşor, pe care l-au rezolvat majoritatea subiecţilor, o valoare a proporţiei apropiată de 0,50 indică un item de valoare mijlocie, iar o valoare apropiată de 0 indică un item dificil sau foarte dificil. Când itemul este prea dificil sau prea uşor, distribuţia acestuia devine asimetrică. Itemii prea uşori discriminează bine între subiecţii slabi, iar itemii dificili între cei buni. Dacă un subiect nu rezolvă un item al cărui indice de dificultate este 0,95 (rezolvat de 95% dintre subiecţii grupului din care face parte) este mai grav decât dacă el ar fi răspuns incorect la un item pe care l-au rezolvat 20% dintre subiecţii grupului de referinţă. Itemii cu indice de dificultate scăzut, deci itemii dificili ($p_i \approx 0,1$), au o asimetrie pozitivă, itemii cu dificultate medie ($p_i \approx 0,5$) sunt simetrici, iar itemii foarte uşori ($p_i \approx 0,90$) au o asimetrie negativă. Mai ales în cadrul probelor de randament, se recomandă ca itemii foarte dificili, pe care nu-i rezolvă aproape nici un subiect, şi itemii foarte uşori, pe care-i rezolvă aproape toţi subiecţii, să fie înlocuiţi. J. Guthke, H.R. Böttcher şi L. Sprung (1991, p. 230) fixează limita inferioară de acceptare a unui item într-un test de randament la 0,20 şi limita superioară la 0,80. Pentru

acelaşi fel de teste, alţi autori indică proporţii care oscilează uşor în jurul valorilor indicate anterior. Aceste limite sunt practicate în teoria clasică a testelor psihologice, în cea probabilistă fiind integraţi în teste şi itemii rezolvaţi corect de un număr foarte mic sau foarte mare de subiecţi. În testele construite după optica probabilistă sunt admişi itemi care pot fi rezolvaţi, de exemplu, de numai de 0,01% dintre subiecţi sau nerezolvaţi doar de acest procentaj de subiecţi. La probele de personalitate, fixarea unor limite de admitere devine mult mai problematică. Pentru aceste probe, numele de *indice de dificultate* este inadecvat; un termen mai potrivit ar fi cel de *indice de proeminenţă*, deoarece prin el s-ar evidenţia gradientul reliefării unui anumit comportament rezultat din operaţionalizarea unei trăsături de personalitate, de exemplu un comportament tipic introvertit.

Opticii clasice de analiză a testului i se reproşează relativitatea proprietăţilor metrice ale itemului. Valoarea indicelui de dificultate calculat depinde de particularitatea eşantionului de subiecţi utilizat în cercetarea privind verificarea calităţilor testului. Astfel, dacă subiecţii sunt slabi, itemul va fi considerat dificil. Din contra, dacă subiecţii au un nivel de competenţă ridicat, itemul va fi considerat uşor. Din acest motiv, reprezentanţii opticii probabiliste de construcţie a testului reproşează „clasicilor" că aprecierea subiecţilor depinde de itemi, iar aprecierea itemilor depinde de subiecţi.

Oferim în continuare un exemplu „cu valori fictive, cuprinzând rezultatele înregistrate într-o *matrice Stern*, la un test cu 10 itemi, administrat unui număr de 11 subiecţi.

Tabelul 17

Matrice Stern cu rezultate binare ale itemilor

Nr. crt.	**Subiect**	**Itemi**										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
1	A.B.	1	1	0	1	1	1	1	1	0	1	8
2	T.R.	0	1	0	1	0	1	0	1	0	0	4
3	M.V.	0	0	1	1	1	1	0	0	1	1	6
4	Z.D.	0	1	0	0	0	1	1	1	0	1	5
5	A.T.	0	0	1	1	0	1	1	1	1	0	6
6	N.F.	0	1	1	1	1	1	1	1	0	1	8
7	U.J.	0	0	1	1	1	1	0	0	0	0	4
8	L.E.	0	1	0	0	0	1	0	0	0	1	3
9	U.I.	0	1	0	1	0	0	1	0	0	0	3
10	D.R.	0	0	0	1	0	1	0	1	0	1	4
11	T.A.	1	1	1	0	0	0	1	0	1	1	6
f_i		2	7	5	8	4	9	6	6	3	7	
p_i		0,18	0,64	0,45	0,73	0,36	0,82	0,55	0,55	0,27	0,64	
q_i		0,82	0,36	0,55	0,27	0,64	0,18	0,45	0,45	0,73	0,36	
$s_i^2 = p_i \times q_i$		0,15	0,23	0,25	0,20	0,23	0,15	0,25	0,25	0,20	0,23	

În partea inferioară a tabelului sunt prezentate, în linie, câteva rubrici rezumative. Semnificaţia acestora este următoarea:

- f_i: suma rezultatelor la un anumit item;

- p_i: proporţia subiecţilor care au rezolvat corect itemul sau, pentru chestionarul de personalitate, proporţia celor care prin răspunsul lor au pus în evidenţă trăsătura investigată;
- q_i: proporţia subiecţilor care nu au rezolvat corect itemul sau, pentru chestionarul de personalitate, proporţia celor care prin răspunsul lor nu au pus în evidenţă trăsătura investigată: $q_i = 1 - p_i$;
- s_i^2: varianţa itemului sau dispersia itemului, care poate fi considerată un indicator grosier al capacităţii de discriminare a itemului. Formula varianţei itemului este: $s_i^2 = p_i \times q_i$.

Poate fi calculat şi indicatorul de asimetrie a itemului prin:

$$\alpha_3 = \frac{q - p}{\sqrt{pq}},$$

unde α_3 este indicatorul de asimetrie pentru item.

Dacă un item are $p = 0,6$ şi $q = 0,4$, se obţine următorul rezultat pentru coeficientul de asimetrie:

$$\alpha_3 = \frac{0,4 - 0,6}{\sqrt{0,4 \times 0,6}} = \frac{-0,2}{0,49} = -0,41$$

Itemul analizat are o asimetrie negativă (valoarea normală pentru asimetrie este 0) şi se poate spune că itemul diferenţiază mai bine subiecţii slabi.

De asemenea, calculul excesului unui item cu exprimare numerică binară constituie o formă particulară pentru calcularea excesului în general:

$$\alpha_4 = \frac{1}{pq} - 6,$$

α_4 reprezentând *indicele de exces sau indicele de kurtosis* al unui item cu exprimare binară a rezultatului.

Aplicând datele din exemplul anterior, obţinem:

$$\alpha_4 = \frac{1}{0,6 \times 0,4} - 6 = \frac{1}{0,24} - 6 = 4,16 - 6 = -1,84.$$

Ţinând cont de faptul că valoarea normală a indicatorului de exces este 3, putem aprecia că acest item are o distribuţie platikurtică.

Prin indicele de dificultate se precizează într-o matrice a punctajelor itemilor vectorul mediei, deoarece este foarte clar că p_i reprezintă şi media distribuţiei pentru valorile itemilor. Această relaţie de calcul nu mai este valabilă la matricele în care valorile itemilor nu mai sunt binare, 0 şi 1. Notaţii diferite pentru itemi apar, de exemplu, în inventarele de interese, în care intensitatea răspunsului este mai fin diferenţiată, în majoritatea cazurilor nedepăşind valoarea 7; deci itemii pot avea valori între 0 şi 7, subiecţii putând obţine oricare dintre valorile 0, 1, 2, ..., 7. Notări diferite de forma binară se întâlnesc şi la probele de randament intelectual, de exemplu WISC, probe la care unele scale de performanţă punctează şi reuşita parţială.

În cazul chestionarelor de interes, cu cât valoarea obţinută la item este mai mare, cu atât subiectul consideră că un anumit interes este mai intens. O astfel de matrice s-ar putea prezenta astfel:

Tabelul 18

Matrice Stern cu rezultate polihotome ale itemilor

Nr.crt.	Subiect	Itemi										
		1	2	3	4	5	6	7	8	9	10	Total
1.	D.B.	1	2	2	2	7	5	3	1	5	1	29
2	L.T.	2	3	1	7	7	4	4	7	4	2	41
3	P.R.	6	4	1	5	5	4	2	7	3	1	38
4	M.D.	3	2	4	6	7	5	4	2	6	1	40
5	F.T.	4	3	3	3	6	4	3	7	7	1	41
6	O.P.	5	5	1	4	7	3	4	4	5	1	39
7	E.T.	2	2	1	2	6	5	6	6	6	2	38
8	D.P.	7	6	1	3	7	4	4	5	4	1	42
9	S.H.	3	4	1	5	6	4	3	4	6	1	37
10	Z.O.	1	3	2	4	5	3	4	3	5	2	32
11	C.I.	1	4	1	2	7	4	3	3	6	1	32
f_i		35	38	18	43	70	45	40	49	57	14	409
p_i		0,45	0,49	0,23	0,56	0,91	0,58	0,52	0,64	0,74	0,18	
q_i		0,55	0,51	0,77	0,44	0,09	0,42	0,48	0,36	0,26	0,82	
$\overline{X_i}$		3,18	3,45	1,64	3,91	6,36	4,09	3,64	4,45	5,18	1,27	
$s_i^2 = \frac{\sum x^2}{N}$, $x = X - \overline{X}$		3,96	1,52	0,96	2,63	0,59	0,45	0,96	4,02	1,24	0,20	

La această formă polihotomică de exprimare a rezultatelor itemului, indicele de dificultate nu se mai confundă cu media itemului şi varianţa itemului nu se mai obţine prin formula expusă anterior. În acest caz, media nu permite să se aprecieze dificultatea relativă a fiecărui item. Pentru calcularea indicelui de dificultate se utilizează formula:

$$p_i = \frac{f_i}{f_i \max},$$

unde p_i este indicele de dificultate, f_i este suma valorilor obţinute la un item de ansamblul subiecţilor (în cazul nostru, 11 subiecţi), f_i max este valoarea maximă pe care ar fi putut-o obţine ansamblul subiecţilor, dacă fiecare ar fi obţinut nota maximă. În exemplul nostru, f_i max este egal cu 77, valoare care rezultă din înmulţirea lui 7 (valoarea maximă pe care o poate obţine un subiect la un item) cu valoarea 11, reprezentând numărul total de subiecţi. Valoarea $p_1 = 0,45$, indicele de dificultate al primului item, se obţine prin împărţirea valorii 35 ($= f_1$) la 77 ($= f_1$ max). Astfel, $p_1 = \frac{35}{77} = 0,45$. Media itemului, notată cu $\overline{X_i}$, se obţine prin împărţirea valorii f_i (suma valorilor unui item) la N (numărul total de subiecţi care au efectuat itemul). Astfel, pentru calculul mediei primului item avem: $\overline{X_1} = \frac{f_1}{N} = \frac{35}{11} = 3,18$, valoare diferită de p_1.

Pentru modul de exprimare polihotomic a rezultatelor unui item, se poate realiza şi o distribuţie pe variante de variaţie, cu exprimarea grafică adecvată.

În unele cazuri, se calculează dificultatea medie a itemilor unui test, notată m_p. În cazul exemplificat de noi,

$$m_p = \frac{\sum_{i=1}^{k} p_i}{k} = \frac{4{,}80}{10} = 0{,}48,$$

unde p_i este indicele de dificultate al itemilor şi k numărul total de itemi, în cazul nostru k = 10. Această valoare este preferabilă mediei testului pentru a exprima tendinţa centrală din cadrul ansamblului itemilor, deoarece media este influenţată de sistemul de notare a răspunsurilor. Indicele de dificultate mediu nu ţine cont de ponderarea individuală a itemilor.

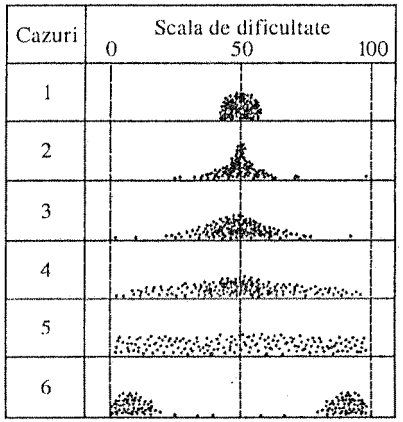


Figura 9. *Reprezentarea grafică a distribuţiei indicatorilor de dificultate ai itemilor unui test*

Indicii de dificultate ai itemilor unui test, diferenţiaţi în clase de dificultate, pot să se distribuie în diverse moduri pe o scală de dificultate:

- distribuţie în formă de punct. Toate sarcinile au dificultatea în jur de 0,50 (cazul 1);
- distribuţie îngust ascuţită a dificultăţii. Majoritatea itemilor au indici de dificultate medii şi numai foarte puţini au indici de dificultate extrem de ridicaţi sau extrem de scăzuţi (cazul 2);
- distribuţie normală a dificultăţii. Distribuţia indicilor de dificultate este apropiată unei distribuţii Gauss-Laplace (cazurile 3 şi 4);
- distribuţie a dificultăţilor aplatizată. Testul conţine aproximativ la fel de mulţi itemi cu dificultate medie şi cu dificultăţi extreme (cazul 5);
- distribuţie a dificultăţii în unghi drept. Itemii foarte dificili sunt la fel de frecvent reprezentaţi ca şi itemii de dificultate medie;
- distribuţie a dificultăţii în formă de U. Numărul de itemi cu dificultate extremă depăşeşte considerabil numărul de itemi medii, în privinţa dificultăţii (cazul 6);

- distribuţie bimodală, în cazul în care două categorii de itemi intermediare ca dificultate sunt la fel de proeminente;
- distribuţie cu neregularităţi. În cazul în care itemi cu dificultăţi diferite formează mai multe vârfuri de frecvenţă.

Valoarea indicelui de dificultate poate fi influenţată de doi factori, şi anume de probabilitatea de reuşită prin hazard şi de numărul de itemi rataţi. Deşi o serie de autori iau în considerare şi itemii rataţi în calculul dificultăţii itemului (*cf.* Laveault şi Grégoire, 1997, p. 225), suntem de părere că analiza de itemi nu poate fi efectuată decât atunci când itemii sunt rezolvaţi în întregime sau când subiecţii ratează foarte puţine răspunsuri. În cazul în care dispozitivul de răspuns este cu alegere multiplă, pot interveni rezolvări prin hazard, subiectul alegând la întâmplare un anumit răspuns, care poate fi cel corect.

În cazul răspunsurilor de tip da/nu, ponderea de reuşită prin hazard poate să atingă 50%, în cazul când există trei posibilităţi de alegere şi doar una este corectă, ponderea de reuşită prin hazard poate să atingă 33%, iar când subiectul are şase posibilităţi de alegere, dintre care doar una este corectă, ponderea de reuşită prin hazard poate atinge 16%.

Se consideră utilă, în unele cazuri, calcularea unei corecţii pentru a preveni deformarea rezultatelor datorită alegerilor corecte prin hazard.

Redăm în continuare un exemplu, cu rezultate fictive, pentru corecţiile practicate în cazul alegerii dintre mai multe variante de răspuns:

Tabelul 19

Corecţii pentru indicatorii de dificultate

	Număr de răspunsuri la alegere			
Subiecţi ↓	**2**	**4**	**6**	**8**
1	1	1	0	1
2	0	1	1	1
3	1	1	1	0
4	1	1	0	1
5	0	0	1	0
6	1	1	0	1
7	0	0	1	1
8	1	0	1	0
9	0	1	0	1
10	1	0	1	1
p_i	0,60	0,60	0,60	0,70
p_i'	0,20	0,47	0,52	0,66
Diferenţa	0,40	0,13	0,08	0,04

Formula pentru corectarea indicelui de dificultate, în cazul existenţei posibilităţii de rezolvare a unor itemi prin hazard, este următoarea (*apud* Laveault şi Grégoire, 1997, p. 226):

$$p_i' = p_i - \left[\frac{1 - p_i}{M - 1}\right]$$

În această formulă, p'_i reprezintă indicele de dificultate corectat; p_i reprezintă indicele de dificultate necorectat; M reprezintă numărul de alegeri la acel item. Această corecţie poate fi realizată doar în unele cazuri, şi nu într-un mod sistematic, deoarece este puţin plauzibilă situaţia ca toţi subiecţii să efectueze alegerea răspunsului corect prin hazard. Corecţia operată nu schimbă ordinea de clasare a subiecţilor.

Vom face o serie de consideraţii asupra varianţei răspunsurilor la itemi, noţiune deosebit de importantă în analizele statistice descriptive şi inferenţiale. Şi în cazul analizei de itemi, varianţa este un indicator al dimensiunii capacităţii de discriminare, care de fapt reprezintă menirea generală a acestui indicator statistic. Dacă un şir de valori ale unei variabile (rezultatele la un test sau la un item) prezintă o valoare constantă pentru toate valorile, atunci valoarea varianţei este egală cu 0 şi, în consecinţă, nu avem nimic de studiat la şirul respectiv. În cercetările psihologice şi în cele care privesc testul psihologic, varianţa joacă un rol deosebit, deoarece în cvasitotalitatea cazurilor ne interesează să opereze discriminări între subiecţi prin intermediul rezultatelor acestora. Este, prin urmare, interesant să se studieze raporturile dintre varianţele rezultatelor la itemi şi varianţa rezultatelor testului în întregime. Vom apela la un exemplu fictiv simplu, care prezintă un număr redus de itemi şi un număr redus de subiecţi.

Tabelul 20

Exemplificare pentru raportul varianţă totală – varianţă a itemilor omogeni

Subiecţi↓	Item 1	Item 2	Item 3	Item 4	Scor X	Item 5	Item 6	X+Item 5	X+ Item 6
1	1	1	1	1	4	1	1	5	5
2	0	0	0	0	0	1	0	1	0
3	0	0	0	0	0	1	0	1	0
4	1	1	1	1	4	1	1	5	5
5	0	0	0	0	0	1	0	1	0
6	0	0	0	0	0	1	0	1	0
7	1	1	1	1	4	1	1	5	5
8	1	1	1	1	4	1	1	5	5
9	0	0	0	0	0	1	0	1	0
Media	0,44	0,44	0,44	0,44	1,77	1	0,44	2,77	2,22
Varianţa	0,278	0,278	0,278	0,278	4,44	0,00	0,278	4,44	6,94

În tabelul prezentat, observăm că itemii 1, 2, 3, 4 sunt rezolvaţi de 4 subiecţi din 9, aceiaşi de fiecare dată (subiecţii 1, 4, 7, 8 răspund corect, iar subiecţii 2, 3, 5, 6 şi 9 răspund greşit). Aceşti itemi au aceeaşi medie şi aceeaşi varianţă şi pot fi consideraţi omogeni, cel puţin din punct de vedere formal. Scorul X constituie scorul compozit al itemilor 1-4. Scorurile compozite X+Item 5 şi X+Item 6 sunt obţinute din suma scorului compozit X şi a scorurilor itemului 5, respectiv 6. Liniile terminale inferioare ale tabelului conţin calculele mediilor şi varianţele scorurilor itemilor şi ale celor compozite. Itemul 5 este reuşit de toţi subiecţii. Scorul compozit (X+Item 5) are, firesc, o medie mai mare cu 1 faţă de cea a scorului compozit X, dar o varianţă identică cu cea a lui X (=4,44). Itemul 6 are aceeaşi medie şi aceeaşi varianţă cu fiecare dintre itemii 1-4 şi, în

mod firesc, media scorului compozit (X+Item 6) este mai mare decât media scorurilor compozite X cu 0,44, dar varianţa sa este superioară cu aproape 2,50. Observăm că varianţa scorului X este mult superioară sumei varianţelor scorurilor de la itemii 1 la 4 (0,278+0,278+0,278+0,278<4,44). De asemenea, observăm că varianţa scorurilor (X+6) este mult mai mare decât suma varianţelor scorurilor implicate (4,44+0,278 < <6,94). Se observă clar faptul că adăugarea unui item omogen (identic ca structură de răspunsuri cu unul dintre itemii existenţi anterior) creşte capacitatea de discriminare a testului.

Pentru a evidenţia efectul schimbării în structura răspunsurilor la itemi, oferim, în continuare, un nou exemplu, în care structurile de răspunsuri sunt eterogene (diferite între ele).

Tabelul 21

Exemplificare pentru raportul dintre varianţa totală a unui test şi varianţa itemilor neomogeni

Subiecţi	Item 1	Item 2	Item 3	Item 4	Scor X	Item 5	Item 6	X+ Item 5	X+ Item 6
1	1	0	1	0	2	1	0	3	2
2	0	1	1	0	2	1	0	3	2
3	0	0	1	0	1	1	0	2	1
4	1	0	0	0	1	1	0	2	1
5	0	1	0	1	2	1	1	3	3
6	1	0	1	0	2	1	0	3	2
7	0	1	0	1	1	1	1	2	2
8	1	0	0	1	2	1	1	3	3
9	0	1	0	1	2	1	1	3	3
Media	0,44	0,44	0,44	0,44	1,766	1,000	0,444	2,766	2,111
Varianţa	0,278	0.278	0,278	0,278	0,250	0,000	0,278	0,250	0,611

Observăm în tabelul de mai sus o situaţie complet schimbată în privinţa varianţei testului, în ciuda faptului că itemii componenţi au aceeaşi medie şi aceeaşi varianţă. Explicaţia acestei stări de fapt o găsim în examinarea atentă a relaţiei care există între itemii componenţi. Astfel, varianţa scorurilor compozite X este mult mai mică decât fiecare dintre varianţele itemilor testului (0,250<0,278). În exemplul anterior, itemii erau eşuaţi sau reuşiţi simultan. În acest ultim tabel, informaţia furnizată de itemi este mai diferenţiată, deoarece avem un ansamblu format din itemi cu structuri (*pattern*-uri) eterogene. Răspunsul (corect sau incorect) dat de către un subiect la un item nu se reproduce la un alt item, considerând în ansamblu situaţia. La o bună parte dintre cazuri, când un item este reuşit, celălalt este ratat şi invers. La testul din primul exemplu, răspunsurile la itemi constituie structuri *omogene*, în timp ce la al doilea test răspunsurile constituie structuri *eterogene*.

Pentru a explica oscilaţiile de valoare ale varianţei totale faţă de varianţele itemilor, vom constitui matricea varianţă – covarianţă pentru itemi binari. Dacă varianţa este un indicator care se referă la o singură serie de valori, covarianţa este un indicator care se

obţine prin prelucrarea a două serii de valori şi exprimă cantitatea comună de variaţie a acestora. În cazul când avem la dispoziţie valorile aceloraşi subiecţi la itemii i şi j, atunci formula covarianţei este:

$$Cov_{ij} = p_{ij} - p_i \times p_j,$$

unde p_i este indicele de dificultate a itemului i, p_j este indicele de dificultate al itemului j şi p_{ij} este proporţia subiecţilor care au rezolvat concomitent itemii i şi j.

Tabelul 22

Cuantificarea coincidenţelor de rezolvare a doi itemi

Nr. crt.	Item i	Item j	Coincidenţă de rezolvare
1	1	0	-
2	1	1	+
3	0	1	-
4	1	1	+
5	0	0	-
6	1	1	+
7	1	0	-
8	1	1	+
9	0	0	-
10	0	1	-
11	1	1	+
	$f_i = 7$	$f_j = 7$	$f_{ij} = 5$
	$p_i = 0,63$	$p_j = 0,63$	$p_{ij} = 0,45$

Aplicând formula de calculare a covarianţei pentru cei doi itemi din tabel, obţinem următorul rezultat: $Cov_{ij} = p_{ij} - p_i \times p_j = 0,45 - (0,63 \times 0,63) = 0,45 - 0,40 = 0,05$.

Pentru a facilita observarea ansamblului combinaţiilor de concomitenţe şi neconcomitenţe ale rezolvărilor corecte, în analiza de itemi se întocmeşte un aşa-numit *tabel al celor patru câmpuri*. Pentru cele două serii de date de mai sus, tabelul celor patru câmpuri se prezintă în felul următor:

		Itemul j		
		Răspuns corect	Răspuns incorect	Total
Itemi	Răspuns corect	a 5	b 2	7
	Răspuns incorect	c 2	d 2	4
Total		7	4	11

Pentru un *tabel al celor patru câmpuri*, explicaţiile pentru conţinutul celulelor sunt:

- câmpul „a" cuprinde numărul de subiecţi care au răspuns corect atât la itemul i, cât şi la itemul j (*pattern* „1-1");

- câmpul „b" cuprinde numărul de subiecţi care au răspuns corect la itemul i şi incorect la itemul j (*pattern* „1-0");
- câmpul „c" cuprinde numărul de subiecţi care au răspuns incorect la itemul i şi corect la itemul j (*pattern* „0-1");
- câmpul „d" cuprinde numărul de subiecţi care au răspuns incorect atât la itemul i, cât şi la itemul j (*pattern* „0-0").
- însumarea câmpurilor „a" şi „b" ne indică numărul de subiecţi care au răspuns corect la itemul i;
- însumarea câmpurilor „c" şi „d" ne indică numărul de subiecţi care au răspuns incorect la itemul i;
- însumarea câmpurilor „a" şi „c" ne indică numărul de subiecţi care au răspuns corect la itemul j;
- însumarea câmpurilor „b" şi „d" ne indică numărul de subiecţi care au răspuns incorect la itemul j.

Pentru un chestionar în care răspunsurile sunt *da* şi *nu* (care pot fi „traduse" numeric în 1 şi 0), Édouard Frienel şi Annick Weil-Barais sugerează posibilitatea existenţei în cadrul unui tabel a unei structuri de date particulare, care fac posibile următoarele tipuri de relaţii între răspunsurile la doi itemi (Weil-Barais, coord., 1997, pp. 200-201):

a) Relaţii de independenţă, în care toate celulele tabelului sunt completate cu numere, diferite de regulă între ele, confirmând existenţa unui anumit *pattern* de răspuns. Itemii sunt consideraţi independenţi, în sensul că măsoară caracteristici diferite.

		Itemul j	
		Da	Nu
Itemul i	Da	X_1	X_2
	Nu	X_3	X_4

În acest tabel găsim prezente *pattern*-urile: Da - Da, Da - Nu, Nu - Da şi Nu - Nu, deci toate tipurile posibile.

b) Relaţii de echivalenţă, în situaţia când toţi subiecţii care au răspuns *Da* la j răspund, de asemenea, *Da* la i, iar toţi subiecţii care răspund *Nu* la j răspund *Nu* şi la i.

		Itemul j	
		Da	Nu
Itemul i	Da	X_1	0
	Nu	0	X_2

Relaţia de echivalenţă presupune că cei doi itemi măsoară aceeaşi realitate.

c) Relaţii de excludere, în situaţia când toţi subiecţii care răspund *Da* la j, răspund *Nu* la i, iar toţi subiecţii care răspund *Nu* la j răspund *Da* la i. O asemenea relaţie presupune că există o excludere reciprocă între caracteristicile măsurate de cei doi itemi.

		Itemul j	
		Da	Nu
Itemul i	Da	0	X_1
	Nu	X_2	0

d) relaţii de ordine sau de implicare, în situaţia când subiecţii care răspund *Da* la i răspund *Da* şi la j, dar nu şi invers, toate celelalte cazuri fiind posibile.

		Itemul j	
		Da	Nu
Itemul i	Da	X_1	0
	Nu	X_2	X_3

În acest caz se prezumă o relaţie de implicare între ceea ce măsoară itemii (i implică j, dar nu şi invers), deci o relaţie de ordine între itemi. O variantă a unei astfel de relaţii este următoarea:

		Itemul j	
		Da	Nu
Itemul i	Da	X_1	X_2
	Nu	0	X_3

În acest caz, j implică i, dar nu şi invers.

Pentru a avea un tablou general al tendinţelor de variaţie a itemilor şi a covariaţiei lor, este util să se trateze *matricea varianţă – covarianţă*. Inspectând această matrice, ne putem face o impresie despre omogenitatea sau eterogenitatea structurii itemilor care compun un test.

Varianţa unui test este egală cu suma varianţelor fiecărui item, la care se adaugă dublul sumei valorilor fiecărei covarianţe posibile în ansamblul itemilor componenţi:

$$s_X^2 = \sum_{i=1}^{n} s_i^2 + 2\sum_{\substack{i,j=1 \\ i<j}}^{n} \text{cov}_{ij},$$

unde s_X^2 este varianţa globală a testului, n este numărul de itemi, s_i^2 este varianţa itemului i şi cov_{ij} este covarianţa itemilor i şi j.

Schema teoretică a matricei varianţă – covarianţă este următoarea (presupunem un ansamblu de şase itemi):

Tabelul 23

Matricea varianţă – covarianţă

			Itemi					
			1	2	3	4	5	6
s^2_T = suma tuturor valorilor	Itemi	1	S^2_1	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}
		2	C_{21}	S^2_2	C_{23}	C_{24}	C_{25}	C_{26}
		3	C_{31}	C_{32}	S^2_3	C_{34}	C_{35}	C_{36}
		4	C_{41}	C_{42}	C_{43}	S^2_4	C_{45}	C_{46}
		5	C_{51}	C_{52}	C_{53}	C_{54}	S^2_5	C_{56}

Pentru cele două tabele din exemplele anterioare, prezentând itemi omogeni şi eterogeni, rezultă, după efectuarea calculelor prin pachetul SPSS, următoarea situaţie:

Tabelul 24

Exemplificare pentru matricea varianţă – covarianţă pentru itemi omogeni şi itemi neomogeni

		Itemi (omogeni)					
		1	2	3	4	5	6
Itemi	1	0,25	0,25	0,25	0,25	0,00	0,25
	2	0,25	0,25	0,25	0,25	0,00	0,25
	3	0,25	0,25	0,25	0,25	0,00	0,25
	4	0,25	0,25	0,25	0,25	0,00	0,25
	5	0,00	0,00	0,00	0,00	0,00	0,00
	6	0,25	0,25	0,25	0,25	0,00	0,25

		Itemi (eterogeni)					
		1	2	3	4	5	6
Itemi	1	0,25	-0,19	0,03	-0,08	0,00	-0,08
	2	-0,19	0,25	-0,08	0,14	0,00	0,14
	3	0,03	-0,08	0,25	-0,19	0,00	-0,19
	4	-0,08	0,14	-0,19	0,25	0,00	0,25
	5	0,00	0,00	0,00	0,00	0,00	0,00
	6	-0,08	0,14	-0,19	0,25	0,00	0,25

Observăm clar în acest din urmă tabel o serie de valori negative care „trag în jos" valoarea de ansamblu a varianţei. Valorile negative exprimă o lipsă de consonanţă a tendinţelor de variaţie. Cu cât se găsesc mai frecvent asemenea valori negative, cu atât

varianţa globală a testului va fi mai scăzută. Inspectarea matricei varianţă – covarianţă ne permite să observăm rapid acele zone de itemi care îngustează împrăştierea valorilor unui test şi, implicit, capacitatea sa de discriminare.

Tabelul din stânga oferă valori mari ale varianţei, de fapt cele mai mari posibile, cu excepţia itemului deosebit de facil, care prezintă un rezultat uniform (doar valori 1) – este vorba despre itemul 5.

Vom face în continuare câteva consideraţii asupra valorilor maximale pe care le pot lua anumiţi indicatori în cadrul operaţiilor efectuate asupra itemilor.

7.2.2. *Indicatori maximali ai itemului*

Varianţa cea mai înaltă posibil a unui item se obţine atunci când media aritmetică (care se confundă, la itemii exprimaţi binar, cu indicele de dificultate) a punctajului itemului este 0,5. Deci valoarea maximală pe care o poate lua varianţa este 0,25, situaţie în care jumătate dintre subiecţi au obţinut scorul 1 şi jumătate scorul 0. Valoarea lui p_i variază în funcţie de calitatea subiecţilor existentă într-un anumit eşantion. Expunem şapte cazuri posibile de evoluţie a valorilor varianţei în funcţie de valorile p_i şi q_i.

Tabelul 25

Valorile varianţei la diferite mărimi ale indicilor de dificultate

Situaţia	a	b	c	d	e	f	g
Valoare p_i	0,00	0,20	0,40	0,50	0,60	0,80	1,00
Valoare q_i	1,00	0,80	0,60	0,50	0,40	0,20	0,00
Varianţa : $p_i \times q_i$	0,00	0,16	0,24	0,25	0,24	0,16	0,00

Se verifică faptul că maximumul varianţei se produce atunci când p_i atinge 0,5, în restul situaţiilor varianţa fiind mai scăzută. Exprimăm grafic aceste situaţii:

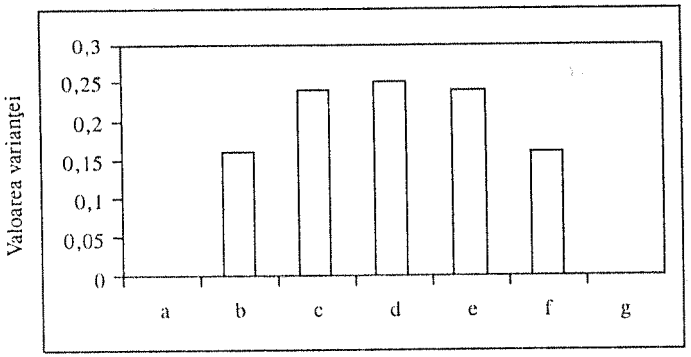


Figura 10. *Histograma valorilor din tabelul 25*

Pentru o înţelegere mai bună a noţiunilor de *varianţă observată a unui test*, *varianţă maximal accesibilă* şi *varianţă maximală*, se prezintă următorul set de date:

Tabelul 26

Date exemplificative pentru studiul varianţelor

Subiect	Exemplul 1				Exemplul 2				Exemplul 3			
	Itemi				Itemi				Itemi			
	1	2	3	X	1	2	3	X	1	2	3	X
1	1	1	0	2	1	0	0	1	1	1	1	3
2	0	1	0	1	0	0	0	0	0	0	0	0
3	1	0	1	2	1	1	1	3	1	1	1	3
4	1	1	0	2	1	1	1	3	1	1	1	3
5	0	1	0	1	0	0	0	0	0	0	0	0
6	0	0	1	1	0	0	0	0	0	0	0	0
7	0	0	1	1	0	0	0	0	0	0	0	0
8	1	0	0	1	1	1	1	3	1	1	1	3
9	1	1	0	2	1	1	1	3	1	1	1	3
10	1	0	0	1	1	1	0	2	1	1	1	3
11	0	0	1	1	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
f_i	6	5	4	15	6	5	4	15	6	6	6	18
p	0,5	0,42	0,33	1,25	0,5	0,42	0,33	1,25	0,5	0,5	0,5	1,5
q	0,5	0,58	0,67		0,5	0,58	0,67		0,5	0,5	0,5	

Exemplele de mai sus constituie situaţii ilustrative pentru varianţa observată, varianţa maximal accesibilă şi varianţa maximală a unui test. Primul exemplu este o situaţie observată, celelalte două – situaţii posibile. Este adevărat că, în anumite conjuncturi, cele trei situaţii se pot confunda, mai ales atunci când omogenitatea unui test este maximă, când itemii au structură omogenă. La primul exemplu, se prezintă situaţia observată în mod real, în care trei itemi au 6, 5 şi 4 răspunsuri corecte şi indicii de dificultate 0,5; 0,42 şi 0,33. Calculul varianţei rezultatului total (rezultatul testului, notat cu X) dă un rezultat egal cu 0,85. Este varianţa observată a unui test. În al doilea exemplu, pentru acelaşi număr de răspunsuri corecte ale celor trei itemi, observăm că varianţa rezultatului total (X) este 1,85. Este varianţa maximal accesibilă, care este mai mare deoarece valorile 1 (accentuate) sunt aliniate, adică itemii sunt omogeni pe cât posibil. În al treilea exemplu se ilustrează situaţia ideală pentru ca varianţa să fie maximă. Varianţa maximală este în acest caz egală cu 2,25. Situaţia este posibilă atunci când, la fiecare item, jumătate dintre răspunsuri sunt corecte, jumătate incorecte, iar răspunsurile 1 sunt perfect aliniate, deci itemii sunt toţi omogeni.

Exemplificările pentru noţiunile de *covarianţă observată* (exemplul 1), *covarianţă maximal accesibilă* (exemplul 2) şi *covarianţă maximală* (exemplul 3) sunt prezentate în tabelul următor:

Tabelul 27

Date exemplificative pentru diverse tipuri de covarianţă (observată, maximal accesibilă, maximală)

Subiect	Exemplul 1		Exemplul 2		Exemplul 3	
	Item i	Item j	Item i	Item j	Item i	Item j
1	1	1	1	1	1	1
2	1	0	1	0	1	1
3	1	1	1	1	1	1
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	1	1	1	1	1	1
8	0	1	0	0	0	0
9	1	0	1	1	1	1
10	0	0	0	0	0	0
11	1	0	1	0	1	1
12	0	0	0	0	0	0
p	0,5	0,33	0,5	0,33	0,5	0,5
	$p_{ij} = 0,25$		$p_{ij} = 0,33$		$p_{ij}=0,5$	
Cov p_{ij}	0,085		0,165		0,250	

În exemplele de mai sus avem trei situaţii, una reală şi două posibile. În primul exemplu, avem situaţia reală a distribuţiei răspunsurilor la doi itemi binari, itemii i şi j, care au dificultăţile 0,5, respectiv 0,33. Covarianţa calculată după formula anunţată este egală cu 0,085. Aceasta este covarianţa observată. În al doilea caz, avem o situaţie posibilă cu doi itemi, i şi j, care au dificultăţile 0,5, respectiv 0,33, deci egale cu cele de la exemplul 1, dar cu o corespondenţă aproape perfectă a răspunsurilor notate cu 1 ; ca atare, p_{ij} nu mai este egal cu 0,25, ca în primul exemplu, ci cu 0,33. Calculul covarianţei produce o valoare mai mare decât la primul exemplu, şi anume 0,165. Aceasta este covarianţa maximal accesibilă. În exemplul al treilea avem structura de reparţie a răspunsurilor la cei doi itemi binari, i şi j, care permite obţinerea celei mai mari covarianţe, situaţia când itemul i este rezolvat de jumătate dintre subiecţi, itemul j este rezolvat de jumătate dintre subiecţi şi ambii itemi sunt rezolvaţi de aceiaşi subiecţi. Este configuraţia din care rezultă cea mai mare covarianţă posibilă, şi anume 0,25. Aceasta este covarianţa maximală.

Covarianţa maximă rezultă din formula enunţată anterior a covarianţei pentru două şiruri de valori binare. Mărimea covarianţei depinde de cota-parte a subiecţilor care au rezolvat corect ambii itemi concomitent. Cota-parte de subiecţi cu două rezolvări corecte concomitente nu poate fi mai mare decât cota-parte a rezolvărilor corecte la itemul cu cea mai mică valoare a lui p din ambii itemi. De aceea, formula poate fi transpusă în felul următor : $C_{max} = p_j - p_i \times p_j$, la care $p_j \times p_i$.

Deci covarianţa maximă este dată de cota-parte a subiecţilor care au rezolvat corect itemul dificil şi care micşorează produsul valorilor celor doi itemi. Dacă, de exemplu, vom avea itemi cu dificultăţile 0,8 şi 0,5, rezultatul maxim va fi 0,10. Prin urmare, covarianţa maximă între doi itemi poate fi exprimată şi în felul următor : $C_{max} = p_j \times q_i$.

În consecinţă, covarianţa maximă este produsul cotelor-părţi ale acelor subiecţi care au rezolvat corect cel mai greu item şi incorect pe cel mai slab. Când ambele cote-părţi sunt egale, rezultă următoarea formulă pentru covarianţa maximă $C_{max} = p \times q$.

Se observă că valoarea maximă a covarianţei este egală cu varianţa maximală a punctajelor itemilor.

Gradul relaţiei dintre doi itemi sau indicele de corelaţie dintre doi itemi poate fi exprimat prin intermediul aşa-numitului *coeficient* φ, a cărui formulă este dată de :

$$\varphi = \frac{p_{ij} - p_i \times p_j}{\sqrt{p_i - p_i^2} \times \sqrt{p_j - p_j^2}}.$$

Toţi termenii formulei sunt explicaţi anterior pentru itemii care prezintă valori binare, 0 şi 1. O formulă similară pentru calcularea coeficientului de corelaţie pentru doi itemi care posedă valori binare este :

$$\varphi = \frac{p_{ij} - p_i \times p_j}{\sqrt{p_i \times q_i} \times \sqrt{p_j \times q_j}}, \text{ deoarece } p_i - p_i^2 = p_i(1 - p_i) = p_i \times q_i.$$

De asemenea, termenii au fost explicaţi anterior.

Matricele de corelaţii sunt necesare în cazul unor analize care urmăresc stabilirea gradului de satisfacere a unor criterii de apreciere pentru anumiţi itemi sau pentru ansamblul acestora.

Exemplificăm calculul unei corelaţii pe datele folosite şi covarianţa.

Nr. crt.	Item i	Item j	Coincidenţă de rezolvare
1	1	0	–
2	1	1	+
3	0	1	–
4	1	1	+
5	0	0	–
6	1	1	+
7	1	0	–
8	1	1	+
9	0	0	–
10	0	1	–
11	1	1	+
	$f_i = 7$	$f_j = 7$	$f_{ij} = 5$
	$p_i = 0,63$	$p_j = 0,63$	$p_{ij} = 0,45$

Astfel, $\varphi = \frac{0,45 - 0,63 \times 0,63}{\sqrt{0,63 - 0,63^2} \times \sqrt{0,63 - 0,63^2}} = \frac{0,45 - 0,40}{\sqrt{0,23} \times \sqrt{0,23}} = \frac{0,05}{0,23} = 0,22$.

Datele organizate într-o *matrice Stern* se pot transforma, prin prelucrările matematice corespunzătoare, într-o matrice de corelaţii. Exemplificăm pentru următoarea matrice cu date binare pentru 6 itemi şi 11 subiecţi. Tabelul din stânga conţine datele primare, cel din dreapta conţine corelaţiile.

Tabelul 28

Matricea de corelaţii interitemi

	Itemi					
Subiecţi ↓	**1**	**2**	**3**	**4**	**5**	**6**
1	1	1	1	1	0	0
2	0	1	0	0	0	1
3	0	0	0	1	1	0
4	1	1	0	0	1	1
5	1	1	0	1	1	0
6	1	0	1	1	0	1
7	0	1	1	1	0	0
8	1	1	1	1	0	1
9	0	0	0	1	0	1
10	1	1	1	0	1	0
11	0	1	0	1	0	0

		Itemi					
		1	**2**	**3**	**4**	**5**	**6**
Itemi	**1**	1,00	0,44	0,15	0,66	0,35	0,77
	2	0,44	1,00	0,66	0,26	0,91	0,44
	3	0,15	0,66	1,00	0,66	0,35	0,77
	4	0,66	0,26	0,66	1,00	0,24	0,44
	5	0,35	0,91	0,35	0,24	1,00	0,35
	6	0,77	0,44	0,77	0,44	0,35	1,00

Ca şi în cazul seriilor de date continue sau discret polihotomice, coeficientul de corelaţie a două serii de date binare poate lua valori pozitive, nule sau negative, cu semnificaţia uzuală. În exemplul dat, nu au rezultat indici negativi de corelaţie datorită particularităţii de structurare a datelor.

Coeficientul de corelaţie joacă un rol important în psihometrie. Calculat prin produsul momentelor (formula lui Karl Pearson), poate varia între –1 şi +1. Aceste limite sunt atinse în cazuri speciale. Pentru testele care produc valori de măsură binare, coeficientul de corelaţie prin momentul produselor este, în general, mai mare strict decât –1 şi mai mic strict decât +1. Corelaţia maximală între doi itemi de aceeaşi valoare rezultă în cazul în care covarianţa între itemii respectivi este maximă.

Pentru aprecierea calităţii itemilor, se efectuează corelaţii între aceştia şi rezultatul global al testului sau, în cadrul operaţiunii de validare, între itemi şi rezultatele unor criterii. Formulele pentru efectuarea corelaţiilor se disting în funcţie de caracterul seriei

binare: dihotomic sau dihotomizat. Seria binară este considerată dihotomică în situaţia în care împărţirea în 0 şi 1 este naturală (de exemplu, răspuns corect/ răspuns incorect) şi dihotomizată când ea traduce o situaţie în care valorile sunt acordate prin convenţie, situaţie întâlnită prin „traducerea" binară a unor date biografice (de exemplu, căsătorit/ necăsătorit, provenienţă din mediul urban/ provenienţă din mediul rural). În primul caz vorbim despre o *corelaţie punct-biserială*, iar în al doilea caz, despre o *corelaţie biserială*. Redăm în continuare două formule de calcul pentru corelaţia punct-biserială. Indicele de corelaţie punct-biserial se notează r_{pbis}.

$$r_{pbis} = \frac{\overline{X}_p - \overline{X}_q}{s_T} \times \sqrt{p \times q} \quad \text{şi} \quad r_{pbis} = \frac{\overline{X}_p - \overline{X}_T}{s_T} \times \sqrt{\frac{p}{q}},$$

unde $\overline{X}_p$ este media acelor valori ale rezultatului total al testului care au valoarea 1 drept corespondent în seria binară a rezultatelor itemilor; $\overline{X}_q$ este media acelor valori ale rezultatului total care au valoarea 0 drept corespondent în seria binară a rezultatelor itemilor; $\overline{X}_T$ este media seriei de rezultate totale; s_T este abaterea standard a seriei de rezultate totale; p şi q sunt cunoscuţi din consideraţiile anterioare, p fiind indicele de dificultate al seriei de date binare, iar q = 1 – p.

Coeficientul de corelaţie biserială are următoarea formulă:

$$r_{bis} = \frac{\overline{X}_p - \overline{X}_T}{s_T} \times \sqrt{\frac{p}{\gamma}},$$

în care singura necunoscută este înălţimea curbei normale în punctul y corespunzând unei densităţi de probabilitate egală cu p, necunoscută notată prin litera γ.

Oferim, în continuare, un exemplu:

Tabelul 29

Date necesare pentru calculul corelaţiei item – rezultat la test

Nr.crt.	**Item**	**...**	**...**	**Rezultat total (X)**	**X_p**	**X_q**
1	1	...	...	37	37	–
2	0			29	–	29
3	1			33	33	–
4	1			42	42	–
5	0			27	–	27
6	0			37	–	37
7	1			32	32	–
8	0			25	–	25
9	1			39	39	–
10	0			32	–	32
11	1			45	45	–
$\Sigma = f_i$	6			378	228	150
p	0,54			$\overline{X}_T = 34{,}36$	$\overline{X}_p = 38$	$\overline{X}_q = 30$
q	0,46			$s_T = 5{,}96$		
p × q	0,248			$s^2_T = 35{,}46$		

În ultimele linii ale tabelului sunt trecute datele necesare în calcularea corelațiilor.

Valoarea γ pentru p = 0,54 este egală cu 0,3958, obținută din tabelul funcției de densitate de probabilitate normală.

Aplicând prima formulă pentru *coeficientul de corelație punct-biserial*, obținem:

$$r_{pbis} = \frac{38-30}{5,96} \times \sqrt{0,54 \times 0,46} = \frac{8}{5,96} \times \sqrt{0,2484} = 1,34 \times 0,498 = 0,668.$$

Aplicând a doua formulă pentru calcularea *coeficientului de corelație punct biserial*, obținem:

$$r_{pbis} = \frac{38-34,36}{5,96} \times \sqrt{\frac{0,54}{0,46}} = \frac{3,64}{5,96} \times \sqrt{1,17} = 0,61 \times 1,08 = 0,658.$$

Observăm că prin cele două formule se obțin practic rezultate identice (jocul aproximărilor face să existe o diferență de 0,01 între cele două rezultate).

Aplicând la aceleași date formula pentru variabile dihotomizate, obținem:

$$r_{bis} = \frac{38-34,36}{5,96} \times \sqrt{\frac{0,54}{0,3958}} = \frac{3,64}{5,96} \times \sqrt{1,36} = 0,61 \times 1,17 = 0,71.$$

Observăm că aplicarea acestei formule produce o ușoară diferență între rezultate.

Toate procedeele de corelare prezentate până în prezent sunt variante ale coeficientului de corelație r introdus de Karl Pearson. În 1968, F.M. Lord și M.R. Novick au demonstrat că un coeficient biserial obținut este cu 20% superior coeficientului de corelație punct-biserial. Următoarea ecuație permite transformarea unei corelații punct-biseriale în corelație biserială:

$$r_{bis} = \frac{\sqrt{pq}}{Y} \times r_{pbis}.$$

În cazul valorilor extreme ale lui p sau q, David Magnusson a demonstrat că o corelație biserială poate fi până la de patru ori mai mare decât corelația punct-biserială, deoarece varianța itemului poate afecta considerabil valoarea maximă pe care o poate lua corelația punct-biserială, echivalent algebric al *r*-ului lui Pearson. Din acest motiv este important, atunci când folosim un program de calcul computerizat, conceput pentru a calcula indicele de discriminare, să știm ce tip de corelație este folosit pentru a calcula corelația item – rezultat total.

7.2.3. *Analiza capacității de discriminare a unui item*

Am amintit faptul că varianța este un indicator grosier al capacității de discriminare a unui item. Această capacitate devine importantă atunci când dorim să analizăm itemii din următorul punct de vedere: un item este apreciat ca discriminativ dacă este rezolvat de o mare proporție dintre subiecții care au obținut un scor ridicat la test în ansamblu, și nu de către subiecții care au obținut un scor slab la test. Totodată, un astfel de item va prezenta o corelație ridicată cu scorul de ansamblu al testului.

Am prezentat mai sus modul de realizare a corelației biseriale și a celei punct-biseriale, cu ajutorul cărora se poate stabili puterea discriminatorie a unui test, urmând logica generală a corelațiilor.

La testele destinate măsurării criteriale, cerința de discriminare nu figurează pe primul plan. În aceste situații, are importanță dacă se depășește sau nu valoarea-prag a unui criteriu, stabilit înaintea procesului de măsurare propriu-zis. Măsurarea criterială se utilizează cu prioritate în evaluarea formativă.

Un indice de discriminare cunoscut și frecvent folosit în practica psihometrică este *indicele de discriminare D*, pus în circulație de W.G. Findley, în anul 1956, realizând practic diferența dintre indicele de dificultate al unui item pentru grupa de subiecți considerată „puternică" după rezultatele globale ale testului (notată p_+) și indicele de dificultate pentru grupa considerată „slabă" (notată p_-). Indicele D poate lua orice valoare între –1 și + 1. Valoarea 0 arată că un item poate fi rezolvat la fel de bine de către un subiect cu un scor total ridicat și de către un subiect cu un scor total scăzut. O valoare negativă arată că un item a fost reușit de o proporție mai mare de subiecți cu un scor total slab la ansamblul testului. Astfel de valori, cuprinse între –1 și 0, ridică semne de întrebare asupra oportunității utilizării unui item. Într-un studiu din anul 1939, privind selecția grupurilor slabe și puternice în operațiunea de alegere a itemilor, T.L. Kelley consideră că grupul puternic este format din acei subiecți cu un scor total care-i plasează valoric în primii 27% dintre subiecți, iar grupul slab din subiecții care au cele mai slabe 27% din ansamblul rezultatelor globale de la test[1]. Să presupunem că s-au examinat 30 de subiecți. Rezultă că grupul puternic cuprinde 8 subiecți (ușoară aproximare a procentului primilor 27% dintre subiecți) iar grupul slab tot 8 subiecți. Oferim un exemplu fictiv cu rezultatele totale ale testului și rezultatul unui anumit item pentru primii și ultimii 8 subiecți.

Nr.crt.	Rezultat test	Rezultat item
1	65	1
2	59	1
3	57	1
4	55	1
5	53	1
6	50	0
7	49	1
8	47	0
...	...	...
23	24	0
24	23	1
25	21	0
26	20	1
27	19	0
28	17	0
29	16	0
30	14	0

1. Acest procentaj este stabilit în articolul lui T.L. Kelley din 1939, „Selection of upper and lower groups for the validation of test items", publicat în *Journal of Educational Psychology*.

Aplicând formulele adecvate obţinem:

$$p_+ = \frac{1+1+1+1+1+0+1+0}{8} = \frac{6}{8} = 0{,}75$$

$$p_- = \frac{0+1+0+1+0+0+0+0}{8} = \frac{2}{8} = 0{,}25$$

$$D = 0{,}75 - 0{,}25 = 0{,}50.$$

R.L. Ebel a propus, în 1965, o scală de repere pentru interpretarea valorică a indicelui de discriminare D:

Interval valoric al indicelui D	Interpretarea discriminării
0,40 şi mai mult	discriminare foarte bună
De la 0,30 la 0,39	discriminare bună
De la 0,20 la 0,29	discriminare slabă
De la 0,10 la 0,19	discriminare de limită
Sub 0,10	fără utilitate

Rezultatul obţinut la exemplul nostru este foarte bun din punctul de vedere al reperelor propuse de Ebel.

Gustav Lienert face o prezentare schematică a raportului dintre discriminare şi dificultatea sarcinilor testelor de inteligenţă. Criteriul pentru discriminare este coeficientul de inteligenţă. Schema este reprezentată în figura următoare (explicaţiile sunt date în tabelul ataşat figurii) (*apud* Lienert, 1967, p. 128).

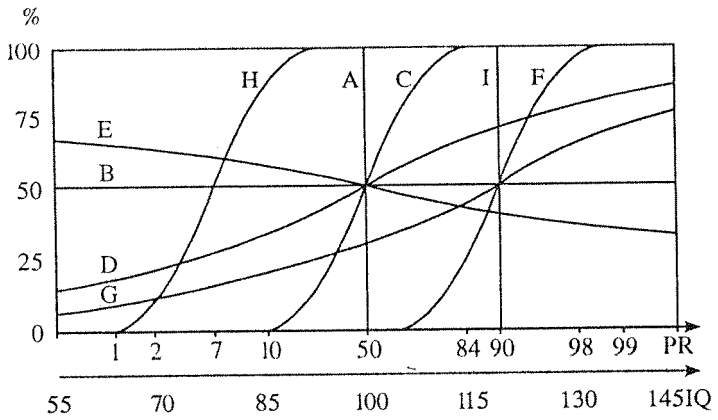


Figura 11. *Reprezentarea spaţială pentru diverse niveluri de discriminare*

A = discriminare absolută	$r_{it} = 1{,}0$	la P = 50
B = fără discriminare	$r_{it} = 0{,}0$	
C = discriminare înaltă	$r_{it} = 0{,}8$	
D = discriminare scăzută	$r_{it} = 0{,}3$	
E = discriminare negativă	$r_{it} = -0{,}1$	
F = discriminare înaltă	$r_{it} = 0{,}8$	la P = 10
G = discriminare scăzută	$r_{it} = 0{,}3$	
H = discriminare înaltă	$r_{it} = 0{,}8$	la P = 93
I = discriminare absolută	$r_{it} = 1{,}0$	la P = 10

R. Crocker şi J. Algina elaborează cinci recomandări pentru alegerea indicilor de discriminare calculaţi pentru itemi dihotomici, comparând metodele corelaţionale cu cea a lui Findley (*apud* Laveault şi Grégoire, 1997, p. 240):

- când itemii sunt de dificultate moderată, ansamblul metodelor se echivalează. Metodele corelaţionale au avantajul de a permite un test de semnificaţie statistică. Un astfel de test nu există pentru indicele D;
- când obiectivul este de a alege itemii situaţi la fiecare extremitate a spectrului nivelurilor de dificultate, se utilizează corelaţia biserială;
- dacă se suspectează că viitoarele eşantioane de subiecţi, cărora le vor fi administrate testele, vor fi puternic diferite de subiecţii pe care a fost aplicat testul, este preferabil să se utilizeze corelaţia biserială;
- dacă se aşteaptă ca testul să fie utilizat pe subiecţi de acelaşi nivel de abilitate, corelaţia punct-biserială este mai indicată;
- când itemul şi variabila-criteriu vor fi cotate în mod dihotomic (scorul total se transformă în „reţinut"/„nereţinut"), trebuie folosit coeficientul tetrachoric, mai ales dacă itemul şi criteriul iau valori extreme. Totodată, deoarece calculul acestei valori este foarte dificil, se preferă coeficientul ϕ.

7.2.4. *Raportul dintre dificultatea şi discriminarea itemului*

Analiza acestui raport este deosebit de importantă în selecţia itemilor unui test aflat în faza de construcţie. Oferim un exemplu privind procedura de urmat atunci când trebuie să analizăm atât dificultatea, cât şi discriminarea.

Să presupunem că examinăm 36 de subiecţi cu un test de 40 de itemi, fără a prezenta analiza decât a primilor 14. În tabel s-au selectat primii 27% dintre subiecţi (care se presupune că au obţinut cele mai bune note la scorul total al testului) şi ultimii 27% dintre subiecţi (care au obţinut cele mai slabe rezultate la acest scor), reprezentând în valoare absolută 10 subiecţi (cu o uşoară aproximare), deci s-a realizat ordonarea subiecţilor în funcţie de mărimea rezultatului global la test.

Tabelul 30

Exemplificare pentru datele necesare în calculul indicelui Findley

Subiecţi →	Itemi													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	1	1	1	1	0	0	1	1	0	0	0	1	0
2	1	1	0	1	1	1	1	0	1	0	0	1	1	0
3	1	0	0	0	1	1	1	0	0	1	0	1	1	0
4	1	0	0	0	1	1	1	0	0	1	1	1	0	0
5	1	0	0	1	1	1	0	1	0	1	0	0	1	1
6	1	0	0	1	1	0	0	1	1	1	1	0	1	1
7	0	1	1	1	1	1	1	1	1	1	0	0	0	0
8	1	0	0	1	0	1	1	0	1	1	0	0	1	0
9	1	0	0	1	1	1	0	0	1	0	0	0	1	0
10	0	0	0	1	1	0	1	1	1	0	0	0	1	0
...														
27	1	1	1	0	1	0	0	0	0	0	0	1	1	0
28	0	1	0	0	1	0	1	0	0	0	1	1	0	1
29	1	0	0	0	1	1	0	0	0	1	0	1	0	0
30	1	0	1	0	0	0	0	1	0	1	1	0	0	0
31	0	1	1	1	0	0	0	0	1	0	0	1	1	0
32	0	1	1	0	1	0	1	1	0	0	1	1	1	1
33	1	1	1	0	1	0	1	0	0	0	1	1	1	1
34	1	1	1	0	1	1	0	0	0	0	0	1	0	0
35	1	1	1	0	1	0	0	0	0	0	0	1	0	0
36	1	0	1	0	1	0	0	0	0	0	0	1	0	0
p_+	0,8	0,3	0,2	0,8	0,9	0,7	0,6	0,5	0,7	0,6	0,2	0,3	0,8	0,2
p_-	0,7	0,7	0,8	0,1	0,8	0,2	0,3	0,2	0,1	0,2	0,4	0,9	0,4	0,3
D	0,1	-0,4	-0,6	0,7	0,1	0,5	0,3	0,3	0,6	0,4	-0,2	-0,3	0,4	-0,1
p_i	0,5	0,5	0,5	0,5	0,9	0,4	0,4	0,4	0,4	0,7	0,1	0,3	0,4	0,2

Cu p_+ şi cu p_- s-au notat indicii de dificultate ai itemilor corespunzători primelor 27% (având rezultatele totale cele mai bune) şi ultimelor 27% (având rezultatele totale cele mai slabe) dintre rezultatele scorurilor la test, cu D s-a notat indicele de dificultate al lui Findley şi cu p_i indicele de dificultate pentru toţi cei 36 de subiecţi. Numărul scăzut de itemi ai unui test duce la mărirea artificială a indicelui de discriminare, deoarece fiecare item are o importanţă prea mare pentru scorul total. O altă exigenţă este cea în privinţa numărului de subiecţi, care trebuie să depăşească o valoare critică de 30.

Curba caracteristică a unui item cu discriminare perfectă se prezintă astfel:

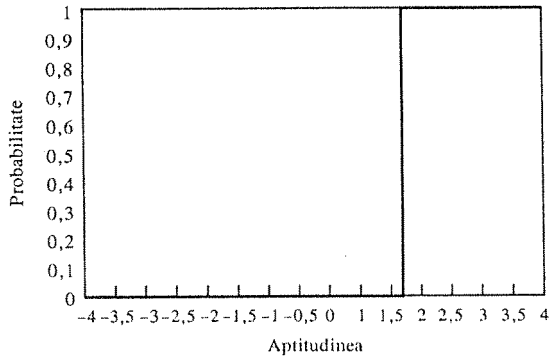


Figura 12. *Curba caracteristică a unui item cu discriminare perfectă*

Prezentăm în continuare un tabel în care sunt ordonaţi indicii de discriminare şi indicii de dificultate ai fiecărui item cuprins în analiză:

Tabelul 31

Date pentru alegerea itemilor adecvaţi

Indice de discriminare	Indice de dificultate								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,9									
0,8									
0,7					4				
0,6				9					
0,5				6					
0,4				13			10		
0,3				7; 8					
0,2									
0,1					1				5
0,0									
– 0,1									
– 0,2	11	14							
– 0,3									
– 0,4					2				
– 0,5									
– 0,6			12		3				
– 0,7									
– 0,8									
– 0,9									

Tipul de indicator de discriminare folosit are mai mică importanţă când itemul este foarte uşor sau foarte greu, deoarece devine riscantă aprecierea contribuţiei sale la diferenţierea subiecţilor la nivelul scorului total. Atât indicele de discriminare D, cât şi indicele de corelaţie sunt într-adevăr influenţaţi de dificultatea itemului.

Combinaţiile diferite de dificultate şi de discriminare au următoarele interpretări:

- itemii 4, 9, 6 şi 13 trebuie să fie reţinuţi, ei reprezentând itemi foarte indicaţi pentru a diferenţia subiecţii: dificultate medie şi discriminare ridicată;
- itemii 7 şi 8 pot să fie reţinuţi. Ei au o capacitate de discriminare la limită şi o dificultate medie;
- itemul 10 merită, de asemenea, să fie reţinut. El discriminează bine şi are un indice de dificultate satisfăcător (la limita de facilitate);
- itemii 1 şi 5 nu servesc discriminării subiecţilor după reperele lui Ebel şi, ca atare, trebuie înlocuiţi. În timp ce itemul 1 are o dificultate medie, itemul 5 este foarte uşor. Dacă am lua drept reper indicele de dificultate, din care am putea deduce varianţa, itemul 1 are o capacitate de discriminare foarte bună, dar itemii reuşiţi sunt prezenţi atât în zona scorurilor bune, cât şi în zona scorurilor slabe. Acesta este motivul pentru care am considerat varianţa ca fiind un indicator grosier al capacităţii de discriminare. La acest item (1) corespund pentru valorile 0 scoruri medii ale rezultatului total;
- itemii 11 şi 14 trebuie retraşi din test, deoarece discriminează negativ. Sunt itemi dificili, reuşiţi mai bine de subiecţii care au rezultate slabe la test;
- itemii 2, 3 şi 12 trebuie să fie şi ei retraşi din test, deoarece au o discriminare negativă accentuată, fiind reuşiţi mai mult de subiecţii din grupul slab. Se poate cerceta în aceste cazuri o eroare în grila de corecţie sau modul defectuos al subiecţilor de a înţelege solicitările itemilor.

Prin sistemul pus la punct de R.L. Ebel se permite articularea indicilor de dificultate şi a celor de discriminare în analiza itemilor unui test. Acelaşi tip de analiză poate fi realizat cu indicii corelaţionali (corelaţia dintre rezultatul testului şi rezultatul unui item, deci corelaţia punct-biserială). În acest caz, valorile maxime de corelaţie se schimbă în funcţie de discriminare şi un test de semnificaţie asupra valorilor acestor corelaţii înlocuieşte pragurile practice propuse de Ebel.

Paul Horst (1971, p. 463) reproşează unui astfel de procedeu că elimină o parte a eşantionului considerat. La acest procedeu de discriminare se iau în considerare doar rezultatele a 54% dintre subiecţii examinaţi, 27% reprezentând rezultatele celor mai buni subiecţi, iar 27% rezultatele celor mai slabi subiecţi. Datorită folosirii îndelungate, metoda este apreciată ca procedeu de analiză acceptat. Dar Paul Horst consideră că fundamentele acestei metode lasă de dorit şi frecvenţa cu care este aplicată poate cu greu să o îndreptăţească. O cantitate mare de informaţii se pierde în situaţia în care, în medie, un procentaj de 46% din eşantion nu este luat în considerare. În consecinţă, autorul amintit nu recomandă un asemenea procedeu, cu toată popularitatea sa. D. Laveault şi J. Grégoire (1997, p. 232), în schimb, îl recomandă pentru analiza itemilor testelor şcolare, cu respectarea condiţiei ca itemii să fie suficient de numeroşi (30 şi mai mulţi). Dacă numărul de itemi este redus, indicele D al lui Findley este supradimensionat, deoarece fiecare item contribuie într-o proporţie prea mare la scorul total.

Testul de semnificaţie a indicilor corelaţionali de discriminare. Dacă un indice corelaţional de discriminare este slab, este important să se stabilească dacă această corelaţie, între item şi scorul total, este semnificativ diferită de 0 sau dacă ea a fost obţinută la întâmplare. Cu cât un eşantion este mai mic, cu atât mai mare va trebui să fie corelaţia, încât să poată fi declarată semnificativ diferită de 0. Cu cât numărul de respondenţi la un test este mai mic, cu atât indicele de discriminare trebuie să fie mai mare, pentru a se putea considera că un item va contribui la diferenţierea subiecţilor în privinţa scorului lor total.

Abaterea standard este utilizată pentru a determina gradul de semnificaţie a corelaţiilor punct-biseriale şi ϕ. În cazul corelaţiei biseriale, abaterea standard a distribuţiei este furnizată prin formula lui A.K. Kurtz şi S.T. Mayo:

$$s_{r_{bis}} = \frac{\sqrt{pq/(N-1)}}{Y},$$

unde Y reprezintă valoarea ordonatei curbei normale a punctului z corespunzător unei densităţi de probabilitate p; p este proporţia de subiecţi care au reuşit itemul; q este proporţia de subiecţi care nu au reuşit itemul; N reprezintă numărul total al cuplurilor de rezultate. Această abatere standard serveşte pentru a determina un interval de încredere de 95% sau 99% în jurul mediei 0. Dacă indicele de corelaţie se situează în interiorul acestor valori, se poate spune că nu există o diferenţă semnificativă de 0 la pragul de semnificaţie de 0,05 şi, respectiv, de 0,01.

7.2.5. *Particularităţi ale indicilor de dificultate şi de discriminare*

La examenele de randament şcolar nu este întotdeauna necesar să se folosească întregul arsenal de instrumente statistice aflat la dispoziţia noastră, în afara analizei de itemi. Această analiză reprezintă un efort ştiinţific considerabil, având în vedere faptul că trebuie să înregistrăm rezultatele tuturor itemilor pentru fiecare subiect. Putem avea eficacitate dacă ştim să exploatăm raporturile care există între principalii indicatori statistici. Proprietăţile indicilor de discriminare se pot utiliza pentru a estima abaterea standard a rezultatelor şi coerenţa internă a unui test. Cu cât este mai ridicată discriminarea medie a itemilor, cu atât ne putem aştepta la o mai mare dispersie a rezultatelor şi la o corelaţie mai puternică între itemi.

W. Wiersma şi S.G. Jurs au demonstrat că suma indicilor de discriminare este de aproximativ 2,45 ori mai mare decât abaterea standard a scorurilor totale. Se poate deci estima abaterea standard prin ecuaţia următoare:

$$s_X = \frac{\sum D}{2{,}45},$$

unde D este indicele de discriminare, iar s_X este abaterea standard a scorurilor totale.

În aceeaşi manieră, Wiersma şi Jurs propun următoarele alternative la formulele K-R 20 şi K-R 21:

$$\text{K-R 20} = \frac{j}{j-1}\times\left[1-\frac{6\sum p\times q}{y\times\left(\sum D\right)^2}\right], \quad \text{K-R 21} = \frac{j}{j-1}\times\left[1-\frac{6\bar{X}\left(j-\bar{X}\right)}{j\left(\sum D\right)^2}\right],$$

unde j reprezintă numărul de itemi. Pentru calculul abaterii standard nu avem nevoie decât de suma indicilor de discriminare D. Pentru a calcula K-R 20 şi K-R 21 avem

nevoie de media scorurilor totale ($\overline{X}$). Utilizând formulele precedente, putem realiza o analiză completă a rezultatelor la un examen plecând de la rezultatele a jumătate dintre subiecţi.

Pentru măsurătorile criteriale, situaţia se modifică, deoarece nu se mai pretinde ca un instrument să diferenţieze în mod egal între toţi subiecţii, ci să-i diferenţieze pe cei care ating sau nu un anumit obiectiv, la un prag de reuşită fixat în avans (cum ar fi analiza progreselor făcute în urma unui proces de învăţare). Itemii utili în măsura criterială sunt cei sensibili la învăţare. Dacă învăţarea a fost profitabilă, gradul de dificultate al acestor itemi trebuie să se schimbe considerabil. Când trebuie să ne pronunţăm asupra atingerii unui obiectiv, aceşti itemi trebuie să ne permită să luăm decizii adecvate. Dacă itemii în chestiune provin de la un acelaşi domeniu de itemi, ei trebuie să fie reuşiţi sau eşuaţi concomitent.

Pentru o evidenţiere mai clară a acestui proces, R.C. Cox şi J.S. Vargas au propus, în 1966, un indice de sensibilitate a învăţării. Calcularea lui se face aflând diferenţa dintre nivelul existent după învăţare (p_{post}) şi cel existent înainte de învăţare (p_{pre}):

$$S = p_{post} - p_{pre}$$

Dacă S este ridicat înseamnă că măsura se realizează cu atât mai mult asupra itemilor, permiţând să se măsoare efectul de învăţare, iar dacă S este mai puţin ridicat, cu atât itemul este mai puţin util, deoarece el se focalizează asupra unor chestiuni care erau la fel de bine reuşite atât înainte de învăţare, cât şi după aceasta. În cazul în care un pretest este reuşit de 34% dintre elevi şi un posttest de 59%, atunci indicele de sensibilitate la învăţare S = 0,59 – 0,34 = 0,25. Un astfel de indice arată că itemii discriminează bine între cei care n-au reuşit itemii înainte de învăţare şi cei care i-au reuşit după învăţare. Valoarea negativă a lui S sau o valoare 0 poate fi interpretată în două moduri:

- itemii nu convin, deoarece ei nu se referă la învăţare:
- învăţarea nu are nici un efect asupra reuşitei subiecţilor.

În cazul testelor criteriale (teste a căror măsură trebuie să posede o serie de caracteristici, precum pertinenţa, fidelitatea, absenţa contaminării), R.L. Brennan a propus un indice similar celui elaborat de Findley pentru calculul discriminării itemilor, estimând pragul de dobândire a unei deprinderi sau dexterităţi. Indicele B este similar indicelui D, mai puţin atunci când grupele puternice şi cele slabe sunt înlocuite cu grupe care au atins sau nu un grad de dobândire a unei deprinderi. Formula este următoarea:

$$B = p_M - p_{NM},$$

unde p_M reprezintă indicele de dificultate al itemului pentru cei care au atins pragul de stăpânire, iar p_{NM} reprezintă indicele de dificultate al itemului pentru cei care nu l-au atins. B poate să varieze între –1 şi +1. Un indice negativ arată că itemul nu discriminează în aceeaşi direcţie ca şi alţi itemi ai testului. Un indice pozitiv indică proporţia de subiecţi din grupul „dobândit" care au reuşit mai bine itemul decât în grupul „nedobândit".

Oferim un exemplu în acest sens. Să presupunem că am realizat tabelul celor patru câmpuri pentru itemii reuşiţi şi nereuşiţi şi pentru cei care stăpânesc sarcina sau nu.

Tabelul 32

Date pentru calculul indicelui Brennan

		Test	
		Nedobândit a+c=21	**Dobândit b+d=13**
Test	**Reuşit a+b=16**	a = 7	b = 9
	Eşuat c+d=18	c = 14	d = 4

$$P_M = \frac{b}{b+d} = \frac{9}{9+4} = 0,69$$

$$P_{NM} = \frac{a}{a+c} = \frac{7}{7+14} = 0,33$$

$$B = P_M - P_{NM} = 0,69 - 0,33 = 0,36.$$

7.2.6. *Analiza omogenităţii itemilor*

Problema privind omogenitatea şi eterogenitatea itemilor şi testelor este intens analizată, autori reputaţi în domeniul teoriei testului psihologic pronunţându-se în diverse moduri (Leon Guttman, Gustav Lienert, Paul Horst, Jane Loevinger, Frederik Kuder şi Marion Richardson). Unul dintre aceşti autori, şi anume Paul Horst, consideră că omogenitatea este un concept fundamental al psihometriei, care nu a fost suficient luat în seamă. Faptul că importanţa conceptului de omogenitate nu a fost recunoscută a creat, în opinia autorului menţionat, o mare dezordine în teoria testului psihologic. Problema definirii precise a fost discutată, în anii '50, de către Jane Loevinger şi Leon Guttman.

Omogenitatea este o noţiune care precizează în ce măsură sunt îndeplinite condiţiile de unidimensionalitate şi libertate faţă de erorile de măsură. Într-o scală de măsură a acestei caracteristici, omogenitatea deplină este o extremă (atunci când ambele condiţii sunt îndeplinite). O omogenitate perfectă este un caz pur teoretic şi nici una dintre condiţii nu poate fi îndeplinită în cazul variabilelor psihologice, în practică înregistrându-se permanent diferite grade de omogenitate.

Există diferenţe între autorii de specialitate în abordarea acestei probleme şi, totodată, dispute ştiinţifice în privinţa sensului termenului de *omogen*. Se vorbeşte despre omogenitatea a două teste, despre omogenitatea a doi itemi sau a itemilor în general, despre omogenitatea itemului şi a testului.

Un exemplu particular asupra omogenităţii îl oferă modelul Guttman. Facem referirile mai ales la itemii probelor de randament intelectual cu itemi structuraţi după modelul Guttman, care presupune că într-o serie de itemi există o creştere permanentă a nivelului de dificultate. Astfel, itemul 2 este mai dificil decât itemul 1, itemul 3 mai dificil decât itemul 2 ş.a.m.d. Un ansamblu de itemi structuraţi după modelul Guttman permite ordonarea ierarhică atât a subiecţilor, cât şi a itemilor. În următorul exemplu, *pattern*-urile 1, 3, 4 respectă modelul Guttman, dar *pattern*-urile 2, 5 şi 6 nu îl respectă, deoarece s-a răspuns la itemi mai dificili şi nu s-a răspuns la itemi uşori. Precizăm din nou că realizarea empirică a unui model este imperfectă, distanţările de model putând fi considerate semnificative sau nesemnificative.

	Itemi			
Pattern de răspuns ↓	1	2	3	4
1	1	1	0	0
2	1	0	1	0
3	1	1	1	0
4	1	1	1	1
5	1	1	0	1
6	1	0	0	1

În concepţia lui Leon Guttman, o serie de itemi poate fi considerată omogenă când respectă modelul său. În acest caz, el vorbeşte despre un *simplex* perfect. Având în vedere că modelul poate fi realizat în cvasitotalitatea cazurilor doar cu imperfecţiuni, se poate spune că o serie de itemi este considerată mai omogenă cu cât respectă mai mult exigenţele amintite. *Pattern*-ul Guttman nu este singurul *pattern* de răspuns care se poate imagina. Deci *omogen* desemnează un item conformat la un model teoretic al *pattern*-urilor de răspuns.

Dacă testul este considerat instrument de investigaţie psihologică, putem afirma că *omogen* se referă la proprietatea a două teste de a măsura aceeaşi trăsătură, de a exprima aceeaşi natură a unei caracteristici. Omogenitatea este o calitate a infrastructurilor identic sau foarte apropiat structurate. Într-un test deplin omogen, ar trebui ca fiecare sarcină să verifice exact aceeaşi funcţie, aceeaşi caracteristică a personalităţii, ca oricare altă sarcină, ca întregul test. Când sarcinile individuale sunt ordonate strict crescător după gradul de dificultate, s-ar putea întâlni (ca excepţie) situaţia în care sarcinile grele să fie rezolvate, iar cele uşoare nu. O asemenea situaţie ar putea apărea atunci când sarcinile individuale măsoară ceva diferit şi când intervin erori accidentale, care reduc utilitatea instrumentală a unui test.

Expunem, în continuare, câteva abordări şi proceduri de stabilire a omogenităţii itemilor şi a testelor în calitate de ansambluri formate din itemi.

Jane Loevinger a avut preocupări intense şi de durată în legătură cu acest concept al omogenităţii, la fel cum au fost şi criticile la adresa modului său de abordare, fapt ce a determinat abandonul cercetărilor ei în această direcţie. În cadrul studiilor pe tema respectivă, autoarea amintită dezvoltă un indice de omogenitate H_t, care se bazează pe compararea a trei variante: varianţa observată a testului (V_x), varianţa unui test deplin eterogen cu aceeaşi distribuţie a dificultăţii sarcinilor (V_{het}) şi varianţa unui test deplin omogen cu aceeaşi distribuţie a dificultăţii sarcinilor (V_{hom}).

$$H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}}$$

După cum arată Jane Loevinger, în expunerile sale despre sarcinile omogene, un test eterogen posedă permanent o variabilitate mai redusă decât un test omogen comparabil. Cu cât într-un test sunt cuprinse concomitent mai multe aspecte singulare ale unei caracteristici, de exemplu inteligenţa generală, cu atât se nivelează mai mult diferenţele individuale. În formula anterioară, varianţa testelor deplin eterogene este scăzută la numitor şi numărător, aşa încât pentru coeficienţii V_x/V_{hom} este definită o scală de punct

nul. Coeficientul H_t însuşi poate fi interpretat în sensul unui coeficient de corelaţie. Ca formulă de estimare pentru H_t, Loevinger indică următoarea expresie:

$$H_t = \frac{N(\sum X^2 - \sum X) + \sum N_i^2 - (\sum X)^2}{2N(\sum iN_i - \sum X) + \sum N_i^2 - (\sum X)^2}$$

unde N = număr de subiecţi; X = valoarea brută a unui subiect; i = locul rangului dificultăţii unei sarcini, din n sarcini, în care cea mai uşoară primeşte rangul 1, iar cea mai grea rangul n; N_i = numărul de subiecţi care au răspuns corect la sarcina cu rangul i.

Aplicarea şi logica imanentă a acestei formule sunt exemplificate în continuare. Un test cu 7 sarcini (notate 1, 2, ..., 7) a fost realizat de 5 subiecţi (notaţi A, B, ..., E), cu rezultatele cuprinse în tabel.

Tabelul 33

Date pentru calculul indicelui Loevinger

Nr.crt.	A	B	C	D	E	N_i	N_i^2	$i \times N_i$
1	1	0	1	1	1	4	16	4
2	0	1	1	1	1	4	16	8
3	0	1	0	1	1	3	9	9
4	0	0	1	0	1	2	4	8
5	0	0	0	1	1	2	4	10
6	0	0	0	0	1	1	1	6
7	0	0	0	0	1	1	1	7
X	1	2	3	4	7	$\sum X = 17$	$\sum N_i^2 = 51$	$\sum i \times N_i = 52$
X^2	1	4	9	16	49	$\sum X^2 = 79$		

În acest tabel, 1 reprezintă un item rezolvat corect, iar 0 un item greşit. Linia X conţine sumele pe coloane, linia X^2 conţine pătratele liniei X, coloana N_i conţine suma pe linii, coloana N_i^2 se obţine ridicând la pătrat valorile N_i, iar coloana $i \times N_i$ conţine produsele dintre coloana N_i şi rangul i (reprezentat în rubrica „Nr.crt."). Sumele rezultate din tabel sunt introduse în formulă:

$$H_t = \frac{5(79-17) + 51 - 17^2}{2 \times 5(52-17) + 51 - 17^2} = 0{,}64.$$

Când se observă localizarea valorilor 1 şi 0, este evident că în colţul din dreapta sus sunt introduse aproape numai răspunsuri corecte, iar în stânga jos numai răspunsuri false. Separaţia nu ar fi aşa de evidentă dacă omogenitatea testului ar fi mai mică. Se ajunge în acest caz la expresia lui Loevinger de definire a omogenităţii: dacă toate sarcinile corespund aceleiaşi funcţii psihice, s-ar întâmpla foarte rar ca o sarcină dificilă să fie rezolvată corect în faţa uneia mai uşoare. Practic, situaţia apare în exemplul dat la candidaţii B, C şi D, care prezintă o valoare 0 în faţa uneia sau mai multor valori 1, dacă se parcurg sarcinile, în ordinea rangurilor, de la 1 la 7. Dacă valorile 1 de pe linii ar fi ordonate accidental, s-ar putea accepta H_t – cum s-ar putea ilustra printr-un rezultat de

valoare nulă sau o valoare diferită numai accidental de aceasta. Ia naştere atunci deplina eterogenitate, iar fiecare sarcină ar verifica altceva.

Dacă se verifică H_t pentru câteva teste care par omogene şi se face constatarea remarcabilă că H_t este scăzută, acest fapt se explică în felul următor:

1. H_t produce un indicator interpretabil fără să se limiteze la testele de nivel care sunt pure. Cu cât este mai mare presiunea temporală, cu atât vor fi mai dispuşi subiecţii să omită anumite sarcini, aşa încât H_t va fi subdimensionat pe nedrept.
2. În mod similar, sarcini pretins omogene iau în considerare grade de dificultate diferite ale anumitor funcţii psihice.

În concepţia lui G. Lienert (1967, p. 118), „omogenitatea unei sarcini este gradul mediu, cu care ea intercorelează cu restul sarcinilor din test". Ecuaţia pe care o prezintă autorul este următoarea:

$$R_{i-n} = \frac{r_{i1} + r_{i2} + \cdots + r_{in}}{n-1} = \frac{\Sigma r_{ij}}{n-1}$$, unde i ≠ j şi n = numărul de itemi.

Simbolurile utilizate în formulă sunt cunoscute. Numitorul fracţiei este n – 1, deoarece autocorelaţia sarcinii r_{ii} nu este introdusă. La intercorelaţii înalte sau foarte diferite, omogenitatea unei sarcini este apreciată mai bine prin transformările z (propuse de Fischer) aplicate pentru r_{ij}. După G. Lienert, conceptele de omogenitate şi discriminare nu sunt independente. Pentru un test eterogen, se pretinde ca fiecare sarcină să posede un minimum de omogenitate la un maximum de discriminare şi, cu aceasta, ar trebui să se examineze corespunzător un alt aspect al aceleiaşi caracteristici a personalităţii. Testele omogene ar trebui să conţină sarcini maximal discriminative şi, simultan, maximal omogene.

Media aritmetică a intercorelaţiilor sarcinilor ar trebui să se poată evalua plecând de la egalizarea intercorelaţiilor conform ordinii de mărime, după o relaţie descoperită de Richardson (1936, *apud* Lienert, 1967), din coeficienţii de discriminare, şi dată cu notaţia originală:

$$_{pf}\bar{r}_{ij} = {}_{pbis}\bar{r}^{2}_{it}$$

unde $_{pf}\bar{r}_{ij}$ reprezintă mediile intercorelaţiilor sarcinilor, exprimate într-un coeficient de corelaţii al celor patru câmpuri (coeficienţii Phi), iar $_{pbis}\bar{r}^{2}_{it}$ – pătratul mediei aritmetice a tuturor coeficienţilor de discriminare, exprimaţi în forma coeficienţilor de corelaţie punct-biseriali.

Sarcinile unui test pot fi mai mult sau mai puţin omogene şi, în cazurile de graniţă, eterogene. Omogenitate înseamnă, în acest caz, unitate de conţinut în perspectiva realizării deplinei independenţe dintre sarcinile individuale. Teste omogene sunt, de exemplu, *Frankfurter Wortschatztest*, ca test general de inteligenţă, testele de organizare a formelor pentru aprecierea potenţialului de reprezentare spaţială, testele seriei de numere pentru evaluarea gândirii deductive.

Astfel de teste omogene cuprind, în mod obişnuit, doar caracteristici ale personalităţii strict circumscrise. În mod curent, ele au o înaltă fidelitate şi pot fi finalizate cu puţine sarcini.

Deşi testele omogene examinează numai aptitudini cu totul speciale, ele sunt valide uneori şi pentru criterii mai generale. S-a demonstrat, de exemplu, că testele de aranjare

a formelor pot să aibă valoare psihodiagnostică de măsurare pentru ceea ce se numeşte uzual inteligenţă practică.

Prin eterogenitatea unei sarcini sau a unui item se înţelege structurarea sa multiplă în privinţa conţinutului, cum se întâlneşte la testele care îşi propun să evidenţieze o caracteristică a personalităţii greu de delimitat sau un întreg complex de caracteristici. Din această categorie de teste, fac parte bine cunoscutele teste de inteligenţă generală ale lui David Wechsler, autor care concepe forme multiple de exteriorizare a inteligenţei. Dintre chestionarele de personalitate consacrate, ar fi considerat o astfel de construcţie psihometrică MMPI-ul lui S.R. Hathaway şi J.C. McKinley.

În privinţa raportului dintre dificultate şi omogenitate, facem precizarea că un $r_{ij} = 1$ nu se obţine decât în cazul în care se răspunde identic la toţi itemii şi doar în cazul dificultăţilor egale ale itemilor. Dacă la unii itemi se răspunde corect mai frecvent decât la ceilalţi, nu ar fi posibilă o intercorelaţie de $r_{ij} = 1$. Cu cât sunt mai mari diferenţele de dificultate, cu atât mai slabi sunt coeficienţii de intercorelaţii. Din această observaţie, rezultă că nu doar omogenitatea de conţinut, ci şi diferenţa de dificultate este relevantă pentru nivelul coeficientului de intercorelaţii. Dacă unii itemi, aparent eterogeni în privinţa conţinutului, corelează relativ slab între ei, acest lucru se poate explica numai prin faptul că itemii unor teste omogene pot fi de dificultăţi diferite.

Cele mai bune condiţii de diferenţiere şi de fidelitate iau naştere la sarcinile necorelate (eterogene) de aceeaşi dificultate sau, în mod corespunzător, la sarcini corelate mediu pozitiv (omogene) de diferite dificultăţi.

Se presupune că există un raport strâns între fidelitate şi omogenitate şi este stabilit faptul că omogenitatea unui test reprezintă un factor important în aprecierea diverşilor coeficienţi de fidelitate. Se pune întrebarea dacă nivelul omogenităţii rezultă doar din comparaţia sarcinilor, deoarece unii teoreticieni ai testului psihologic nu indică omogenitatea ca pe o calitate proprie testului, ci o atribuie noţiunii consistenţei interne. În practică, testele cu itemi deplin omogeni sunt şi suficient de consistente, în sensul pe care îl conferă consistenţei formulele Kuder-Richardson. Invers, testele care produc la analiză un înalt coeficient de consistenţă sunt de cele mai multe ori omogene. Acest raport nu este surprinzător, deoarece unele formule Kuder-Richardson au luat naştere din formulele privind corelaţia de sumă, prin postularea că sarcinile elementare (sau itemii) sunt deplin omogene.

Din punctul de vedere al teoriei clasice a testului psihologic, se pot face următoarele consideraţii în legătură cu noţiunea de omogenitate:

- Un test poate fi considerat omogen când toţi itemii lui măsoară în aceeaşi proporţie valoarea adevărată. Când jumătăţile au aceeaşi medie şi aceeaşi dispersie a punctajului, un coeficient de echivalenţă pentru teste omogene este cea mai bună estimare pentru coeficientul de precizie al unui test. Situaţia stă altfel când exprimăm coeficientul de precizie al unui test eterogen printr-un coeficient de echivalenţă. Eterogenitatea exprimă faptul că itemii individuali măsoară aceeaşi valoare adevărată în diverse grade. Dacă itemii sunt ordonaţi în funcţie de gradul de dificultate pe ambele jumătăţi ale testului, există riscul ca sarcinile de un anumit fel să se aglomereze într-o jumătate, iar sarcinile de alt fel în cealaltă. Nu se poate controla cum se grupează itemii de acelaşi fel şi de constituţii diferite în ambele părţi. Testele se structurează în funcţie de media şi dispersia lor în mod egal, nu şi în funcţie de conţinutul lor. Un coeficient de înjumătăţire a testului, aflat prin

înjumătăţirea unui test pe considerentul frecvenţelor de rezolvare, nu constituie o exprimare adecvată pentru fidelitatea testului în semnificarea preciziei. El nu reprezintă nici o bună exprimare pentru consistenţa internă a testului. De la început, jumătăţile testului se construiesc în aşa fel încât condiţiile de paralelism, atât cele referitoare la media şi dispersia distribuţiei datelor, cât şi cele referitoare la conţinutul sarcinilor, să fie îndeplinite.

- Metoda înjumătăţirii testului poate să fie utilizată pentru estimarea preciziei unui test, estimarea empirică realizându-se sub forma unui coeficient de echivalenţă. Pentru teste omogene, acesta poate fi calculat între itemii cu numere de ordine pare şi cei cu numere de ordine impare, dacă itemii testului sunt ordonaţi la întâmplare sau în funcţie de gradul lor de dificultate. Calcularea coeficientului de echivalenţă pentru teste eterogene trebuie să se bazeze pe conţinutul şi pe dificultatea sarcinilor perechi.
- Dacă dorim distingerea omogenităţii de consistenţa internă, este indicat să descriem factorii sarcinilor individuale. În acest fel, s-ar realiza o analiză a tuturor factorilor unui test, la care ar trebui interpretate intercorelaţiile unor itemi şi operată inserarea acestora în matricea factorilor. Un test ar fi cu atât mai omogen cu cât toate sarcinile ar fi mai comune şi cu cât ar fi încărcate mai puţin cu factori specifici. Analiza factorială a itemilor oferă posibilitatea de a forma grupe de itemi relativ omogeni, care pot apoi să funcţioneze ca subteste ale unei baterii de teste şi să constituie un profil de test.

Se vorbeşte de multe ori despre omogenitate la capitolul dedicat analizei fidelităţii, fiind considerată o fidelitate funcţională. Dar este necesar să subliniem ideea potrivit căreia consistenţa semnifică exclusiv precizia instrumentului de măsură, iar omogenitatea semnifică unitatea factorială a funcţiei psihice de măsurat. În acest sens, identificarea cu fidelitatea funcţională nu apare ca suficient de fundamentată. Dacă definim omogenitatea astfel, atunci este fixată relaţia dintre omogenitate şi validitate : omogenitatea este o condiţie a validităţii factoriale. Testele factorial valide trebuie să dovedească o structură factorială unitară şi, în afară de aceasta, să fie interpretabile, ceea ce nu este necesar la precizarea noţiunii de *omogen*.

Paul Horst foloseşte raporturile dintre indicatorii maximali în analiza omogenităţii interitemi, a omogenităţii a două teste şi a omogenităţii item/test. În primul caz, se referă la indicatorii maximali care pot fi stabiliţi pur teoretic, pe baza vectorilor de dificultate întâlniţi la matricele binare. În al doilea rând, se referă la corectura (accidentală) a acestei valori pe baza rezolvărilor întâmplătoare. Este vorba astfel despre conceptul de *indicator maximal* şi de *indicator accesibil maximal*. Se pun în relaţie varianţa, covarianţa şi distribuţia punctajelor la test, valorile punctajelor testului putând fi compuse din valori de măsură binare. La punctajele individuale ale itemilor se poate stabili doar covarianţa maximal accesibilă, deoarece în modul cum sunt stabilite mediile itemului sunt fixate şi varianţele.

Putem să precizăm omogenitatea unui indicator în două feluri, o dată ca raportul dintre indicatorul observat şi cel maximal la mediile stabilite ale itemului, apoi ca raportul dintre indicatorul observat şi indicatorul accesibil maximal. Această valoare accesibilă maximal este o funcţie a valorilor observate şi a parametrului întâmplător. Există argumente pro şi contra cu privire la cele două definiţii ale omogenităţii. Raportul dintre indicatorii observaţi şi indicatorii maximali nu poate să depăşească niciodată 1 şi, în mod

curent, nu poate atinge această valoare din cauza influenţelor întâmplătoare. Respectiva valoare este un reper în practică, deoarece poate fi oportună existenţa unui indice de omogenitate, care să atingă în situaţii practice 1 (se utilizează mai curând raportul dintre indicatorul observat şi indicatorul maximal accesibil). Acest indice de omogenitate are un avantaj : el poate fi considerat o măsură pură pentru omogenitate. Atunci când utilizăm maximumul matematic, în raport se introduce fidelitatea măsurării. Este mai adecvată însă considerarea omogenităţii independent de fidelitate sau de fluctuaţiile întâmplătoare. Dacă un test prezintă itemi cu alegere corect/fals sau cu dublă alegere, unde verosimilitatea de a avea un răspuns corect pur întâmplător este de 0,50, atunci estimarea omogenităţii cu indicatori maximali la numitor produce o valoare mai scăzută decât dacă itemii testului ar avea mai mult de două posibilităţi de răspuns.

Teoretic, este posibil ca raportul dintre indicatorul observat şi indicatorul maximal accesibil să depăşească 1, deoarece valoarea maximal accesibilă, după formula amintită anterior, poate să fie mai mică decât valoarea observată. În cele mai multe situaţii practice este totuşi mai bine ca la numitor să se afle valoarea maximal accesibilă, nu valoarea maximală. Raportul rezultat este cel puţin o graniţă superioară pentru omogenitatea itemilor testului. Două rezultate diferite, care se obţin pe baza diverşilor numitori, au deci limite mai înalte sau mai joase ale omogenităţii.

O critică la propunerea de a prezenta omogenitatea unui indicator printr-un raport este aceea că raportul a fost ales aparent după o opinie, după o părere. S-ar putea să existe alte funcţiuni ale indicatorilor observaţi şi ale celor maximali sau maximali accesibil, având 1 ca margine superioară. Însă acest argument nu este restrictiv. La toate cercetările ştiinţifice şi la construcţiile multor modele teoretice, este uzuală considerarea unor funcţii matematice, care sunt rezonabile pe o bază apriorică. Din numărul posibil de funcţii care satisfac criteriile considerate, ar trebui să se aleagă una dintre funcţiile cele mai simple. Utilizarea rapoartelor pentru estimarea omogenităţii itemilor este îndreptăţită în lipsa unei alternative mai bune.

Omogenitatea a doi itemi dintr-un test. Tratăm în două moduri măsura pentru omogenitate, şi anume : raportul dintre covarianţele observate şi cele maximale şi raportul dintre covarianţele observate şi cele maximal accesibile la două teste binare. Coeficientul de corelaţie Φ pentru doi itemi ai testului de dificultate inegală este o subestimare a raportului dintre ambele funcţii considerate, deoarece corelaţia maximală între doi itemi de dificultate inegală aparţinând testului este totdeauna mai mică decât 1, luată în valoare absolută. Aceasta este o limitare pur matematică, fără a lua în considerare cât de asemănători sunt itemii în privinţa dificultăţii. Loevinger propune ca pentru efectuarea corecturii acestei discrepanţe corelaţia observată să se împartă la corelaţia maximală, operaţie posibilă la doi itemi având grade de dificultate diferite. Se poate arăta uşor că acest raport este independent de varianţele celor doi itemi şi că numai covarianţele observate necesită împărţirea la covarianţele maximale. Propunerea lui Loevinger a fost criticată de J.B. Carroll, care preferă pentru variabilele codificate binar utilizarea coeficientului de corelaţie tetrachoric. Formula coeficientului de corelaţie tetrachoric este :

$$r_{tet} = \cos\frac{\pi\sqrt{b\times c}}{\sqrt{b\times c}+\sqrt{a\times d}},$$

unde a, b, c şi d sunt frecvenţele calculate în tabelul celor patru câmpuri.

Raportul dintre covarianţa observată şi covarianţa maximală. Se consideră că raportul dintre covarianţa observată covarianţa maximală este un indice de omogenitate simplu şi practic pentru doi itemi ai testului (vezi P. Horst, 1971, p. 285). Formula indicelui este:

$$h_{ij} = \frac{p_{ij} - p_i p_j}{p_j(1 - p_i)}.$$

În această formulă, cu h_{ij} s-a notat indicele de omogenitate. Celelalte simboluri utilizate sunt cunoscute. Se consideră că $p_j \leq p_i$. Un alt mod de exprimare a formulei este:

$$h_{ij} = \frac{p_{ij}/p_j - p_i}{1 - p_i}.$$

Un exemplu numeric este dat pentru $p_i = 0,6$, $p_j = 0,4$ şi $p_{ij} = 0,3$. Se obţine:

$$h_{ij} = \frac{0,3 - (0,6)\times(0,4)}{(0,4)\times(1,0 - 0,6)} = \frac{0,06}{0,16} = 0,375.$$

Indicele de omogenitate tratat anterior poate fi numit al „omogenităţii minimale" (în terminologia engleză, *lower limit homogeneity*). Deoarece covarianţa maximal accesibilă are, în general, o valoare mai scăzută decât covarianţa maximală, vrem să desemnăm raportul dintre covarianţa observată şi covarianţa maximal accesibilă ca „omogenitatea maximală" (*upper limit homogeneity*):

$$H_{ij} = \frac{p_{ij} - p_i p_j}{(p_j - c)\times(1 - p_i)},$$

unde c este un parametru al rezolvărilor întâmplătoare, celelalte simboluri fiind cunoscute din consideraţiile anterioare.

Pentru ilustrare, folosim datele anterioare şi $c = 0,125$, valoare care presupune că mecanismul de răspuns are 8 posibilităţi de răspuns.

$$H_{ij} = \frac{0,3 - 0,4\times 0,6}{(0,4 - 0,125)\times(1 - 0,6)} = \frac{0,06}{0,11} = 0,54.$$

Omogenitatea unui test. Reamintim că prin omogenitatea unui test se înţelege proporţia în care itemii testului măsoară acelaşi lucru. Această proporţie este dată prin covarianţele între itemii individuali. Putem să spunem că itemii testului măsoară acelaşi lucru, atunci când covarianţele itemilor (la dificultăţi date ale itemilor) sunt cât mai mari posibil. Deoarece varianţa este maximă în cazul în care covarianţele itemilor la dificultăţi date sunt maxime, omogenitatea unui test se defineşte ca raportul dintre varianţele observate ale testului şi varianţa maximală a testului sau cea maximal accesibilă.

Tratăm, în primul rând, indicele cel mai mic de omogenitate. La metoda aleasă, indicele de omogenitate poate lua valori între 0 şi 1. Ecuaţia pentru definirea indicelui de omogenitate a unui test, ca raportul dintre varianţa observată şi varianţa maximal accesibilă, este următoarea:

$$H_s = \frac{\sigma_S^2}{2\sum_{i=1}^{n} ip_i - \bar{S}(1 + \bar{S})}.$$

Prezentăm un exemplu numeric pentru indicele de omogenitate al unui test ca funcţie a mediei sale, a abaterii standard şi a vectorului dificultăţilor itemilor.

$$\sigma_S^2 = 2,3 \quad \bar{S} = 4,3 \text{ şi } p_{X'} = \{1,0;\ 0,9;\ 0,8;\ 0,6;\ 0,5;\ 0,3;\ 0,2\}$$

$$h_S = \frac{2,3}{2[1,0 + 2(0,9) + 3(0,8) + 4(0,6) + 5(0,5) + 6(0,3) + 7(0,2)] - (4,3)(5,3)} =$$

$$= \frac{2,3}{2(1,0 + 1,8 + 2,4 + 2,4 + 2,5 + 1,8 + 1,4) - (4,3)(5,3)} = \frac{2,3}{26,6 - 22,79} =$$

$$= \frac{2,3}{3,81} = 0,60.$$

Raportul dintre varianţa observată şi varianţa maximal accesibilă este numit indicele cel mai înalt de omogenitate al unui test. Formula sa este următoarea:

$$H_S = \frac{\sigma_B^2}{2(1-c)\sum_{1}^{n} i\times p_i + \bar{S}(2n\times c - 1 - \bar{S}) - c\times n(n-1)}$$

Utilizăm un exemplu numeric cu datele anterioare şi, în plus:

$$\sum_{1}^{n} ip_i = 13,3;\ c = 125;\ n = 7$$

Se obţine:

$$H_S = \frac{2,3}{2(0,875)13,3 + 4,3[(2)7(0,125) - 5,3] - (0,125)7(6)} =$$

$$H_S = \frac{2,3}{23,27 - 15,265 - 5,25} = \frac{2,3}{2,75} = 0,83.$$

Omogenitatea item – test. Măsurarea omogenităţii item – test prezintă un deosebit interes şi ar trebui să producă o valoare de măsură care să indice faptul că un item individual are acelaşi conţinut ca şi ceilalţi itemi din test. În cazul în care alţi itemi din test nu măsoară toţi acelaşi lucru, există posibilitatea ca un anumit item individual să nu cuprindă natural toate aspectele pe care le măsoară ceilalţi itemi. Este, prin urmare, firesc ca indicele de omogenitate item – test să reflecte atât omogenitatea itemilor testului, cât şi concordanţa fiecărui item cu ceea ce cuprind în general itemii testului. Ca atare, valoarea cea mai mică a omogenităţii item – test se defineşte în calitate de covarianţă observată a itemului şi a testului, împărţită la covarianţa maximală pe care poate să o aibă itemul şi un test cu aceleaşi dificultăţi ale sarcinilor (în acest punct se acceptă că itemul nu face parte dintre itemii testului). Se utilizează următoarea formulă:

$$H_{kS} = \frac{C_{kS} - p_k q_k}{\sum_{i=k+1}^{n} p_i + p_k(k - \bar{S}) - p_k q_k}$$

În continuare, un exemplu numeric. Pentru

$p_k = 0,6$; $q_k = 0,4$; $k = 2$; $\bar{S} = 4,3$; $C_{kS} = 0,55$;

$p_{X'} = \{1,0; 0,9; 0,8; 0,6; 0,5; 0,3; 0,2\}$

se obţine cea mai mică omogenitate item - test:

$$h_{kS} = \frac{0,55 - 0,6 \times 0,4}{(0,8 + 0,6 + 0,5 + 0,3 + 0,2) + 0,6(2 - 4,3) - 0,6 \times 0,4} =$$

$$= \frac{0,55 - 0,24}{2,4 - 1,38 - 0,24} = \frac{0,31}{0,78} = 0,39$$

Valoarea cea mai înaltă a omogenităţii item - test, ca raport dintre două covarianţe, este dată prin următoarea ecuaţie:

$$H_{kS} = \frac{C_{ks} - p_k q_k}{(1-c)\left[\sum_{i=k+1}^{n} p_i + p_k(k - \bar{S}) - c(n - 1 - \bar{S}_t)\right] - p_k q_k}.$$

Utilizând datele anterioare, obţinem:

$$S_t = \frac{\bar{S} - n \times c}{1 - c} = \frac{4,3 - 7(0,125)}{0,875} = \frac{3,425}{0,875} = 3,91$$

$$H_{kS} = \frac{0,55 - 0,6 \times 0,4}{0,875[2,4 + 0,6(2 - 3,91) - 0,125(6 - 3,91)] - 0,24} = 0,50$$

Omogenitatea a două teste. Omogenitatea a două teste se defineşte ca fiind raportul dintre covarianţele observate şi/sau covarianţele lor maximale sau dintre covarianţele maximal accesibile şi dificultăţile prestabilite ale itemilor. Se observă că, în această definiţie, se relaţionează nu numai omogenitatea între ambele teste, ci şi omogenitatea itemilor în interiorul fiecărui test.

Dacă omogenitatea a două teste se poate defini în mod firesc prin omogenitatea itemilor corespunzători din ambele teste, aceasta ar însemna ordonarea ansamblului de perechi de itemi din ambele teste. O astfel de împerechere s-ar baza, probabil, pe decizii întâmplătoare şi foarte subiective.

Se preferă, pentru omogenitatea testului, utilizarea raporturilor care produc cele mai mici valori, deci pentru termenul de covarianţă de la numitor se acceptă omogenitatea maximală în interiorul şi între teste.

7.2.7. *Estimarea specificităţii unui test*

Specificitatea şi omogenitatea sunt extreme opuse ale conceptului de covarianţă între variabile. Covarianţa slabă este prezentă la specificitatea itemului, covarianţa înaltă la omogenitatea itemului. Aceste relaţii au valoare în funcţie de itemi sau de teste. Când itemii unui test sunt suficient de omogeni, putem să observăm corelaţia dintre două teste omogene. Când aceasta este scăzută, avem, referitor la variabila măsurată, un înalt grad de specificitate. O corelaţie înaltă reflectă şi o omogenitate înaltă a testului. Când ne referim la testele dintr-o baterie de teste, trebuie luate în considerare toate testele, dacă dorim să indicăm care dintre ele sunt specifice. Să presupunem că într-o baterie de teste găsim unul care corelează scăzut cu toate celelalte. În cazul în care itemii din acest test

sunt omogeni, atunci putem spune că testul posedă un înalt grad de specificitate. Corelaţiile sale cu celelalte teste din baterie sunt în acest caz scăzute. Dacă acceptăm un coeficient înalt de omogenitate pentru testul respectiv, atunci putem spune că el posedă o înaltă specificitate.

Când vrem să facem prognoze cu o baterie de teste, este important, în general, ca testele individuale ale bateriei să posede un înalt grad de specificitate. Conceptele de specificitate şi fidelitate sunt strâns legate unul de celălalt. Semnificaţia acestei relaţii pentru măsurare şi prognoză în psihologie este însă rar reliefată în mod corect. În cazul ideal, pentru o baterie de teste, itemii aceluiaşi test trebuie să fie înalt omogeni, dar când este vorba despre itemii altor teste, ei trebuie să aibă un înalt grad de specificitate. Specificitatea prezintă importanţă în scopul evitării suprapunerii, precum şi în prognoza unui criteriu complex şi prognoza diferenţială a unui criteriu multiplu.

Ca fundament pentru cerinţa ca fiecare test al unei baterii să posede un înalt grad de specificitate, stau aspectele de economie temporală şi financiară. Nu are sens să reunim, într-o baterie de teste, diverse teste care măsoară acelaşi lucru. Aplicarea a două teste care corelează înalt înseamnă risipă de efort şi timp.

În unele cazuri, pare natural şi necesar să se aplice mai multe teste diferite pentru aceeaşi funcţie sau mai multe forme paralele ale aceluiaşi test. Este însă cu totul altceva, când avem doar două teste care cuprind numai aproximativ, dar nu exact acelaşi lucru. Din motive de siguranţă, este important să dispunem de forme paralele ale aceluiaşi test, aşa încât să realizăm diverse forme, în diverse momente şi în locuri diferite, la diverse grupe de subiecţi. Astfel, pot fi reduse la minimum posibilităţile de eroare. Este de preferat să se apeleze la teste care se suprapun pe cât de mult posibil, în scopul de a realiza estimarea fidelităţii unui procedeu sau a altuia. În acest caz, avem de-a face cu teste care trebuie să diagnosticheze aceeaşi variabilă a comportamentului sau acelaşi proces psihic, în acelaşi mod.

Cazul indezirabil al suprapunerii apare atunci când două sau mai multe teste măsoară practic aceleaşi funcţii, însă suprapunerea lor nu este atât de puternică încât să poată servi ca teste paralele.

Un test cu o înaltă specificitate, în interiorul unei serii de alte teste, posedă o înaltă saturaţie într-un factor specific, însă numai saturaţii scăzute în factori generali.

Specificitate şi corelaţie multiplă. Un test este deplin specific atunci când toate corelaţiile sale cu alte teste observate sunt nule. Acest caz nu apare, de regulă, la variabilele experimentale. Când un test este aplicat la un eşantion de subiecţi, în interiorul unei baterii de teste, pe baza corelaţiilor multiple poate fi dedus un index de specificitate. Definirea specificităţii prin intermediul analizei factoriale se bazează pe varianţa în factorul specific. Dacă la bază se află analiza factorială, cu o serie de corelaţii între perechile de variabile, putem defini specificitatea unui test ca funcţie a corelaţiei sale multiple cu toate celelalte teste ale bateriei de teste sau, mai simplu, ca 1 (unu) minus pătratul acestor corelaţii multiple. Valoarea respectivă reprezintă partea de varianţă a testului, care în acest eşantion special nu poate fi estimată pe baza altor teste şi, prin urmare, poate fi apreciată ca specificitate a testului.

Definiţia specificităţii se află în strânsă relaţie cu o anumită tehnică a analizei factoriale, cunoscută sub denumirea de *image analysis* şi dezvoltată de Guttman, în 1953.

Comunalitatea unui test este definită ca fiind acea cotă-parte a varianţei, pe care testul o are în comun cu alte teste măsurând aceeaşi variabilă. Graniţa de jos a acestei

comunalităţi este cunoscută ca pătratul corelaţiilor multiple ale respectivului test cu toate celelalte teste ale bateriei. Aşadar, specificitatea poate fi definită ca „unu" minus comunalitatea.

Specificitatea unui test poate fi tratată cel mai bine atunci când se fondează pe relaţiile unei variabile cu alte variabile. Aceste variabile pot fi rezultate din suma sau din suma ponderată a valorilor de punctaj sau chiar punctaje ale itemilor.

Estimarea din corelaţii. Metoda este rapidă, estimativă şi utilizează pentru calculul specificităţii testului k formula următoare:

$$p_k = 1 - \frac{r_{kh}^2}{r_{kk} r_{hh}},$$

unde r_{kh} este corelaţia testelor k şi h din bateria de teste, testul h fiind acela cu care corelaţia testului k este maximală; r_{kk} şi r_{hh} sunt fidelităţile testelor h, respectiv k.

De exemplu, pentru $r_{kh} = 0{,}32$, $r_{kk} = 0{,}70$ şi $r_{hh} = 0{,}80$ se obţine: .

$$r_k = 1 - \frac{(0{,}32)^2}{(0{,}70)(0{,}8)} = 1 - \frac{0{,}1024}{0{,}56} = 0{,}817,$$

interpretarea fiind cea uzuală, a unei specificităţi înalte.

Capitolul III

ASPECTE ALE CONSTRUCŢIEI TESTELOR

1. Raţiunea construcţiei unui test

În momentul în care un psiholog îşi propune să construiască un test, acesta are, desigur, o motivaţie bine întemeiată pentru intenţia sa, rezultată din neacceptarea variantelor de teste deja existente, care i se oferă pentru atingerea unui obiectiv psihodiagnostic, din cunoaşterea structurii unui test şi a cantităţii de efort competent presupus pentru ducerea la bun sfârşit a proiectului său. Cel mai adesea, un test este pus în serviciul diagnozei transversale sau de profil, prin care se urmăreşte cunoaşterea locului ocupat de un individ în ansamblul unui grup de indivizi comparabili, în legătură cu o anumită caracteristică psihică. Testul psihologic poate stabili, de asemenea, gradul de reliefare a unei anumite caracteristici a personalităţii între indivizi diferiţi şi între diverse grupuri. Prin aplicarea unui test psihologic se poate întemeia decizia privind îndeplinirea sau neîndeplinirea unei condiţii care să certifice prezenţa unei caracteristici pretinse pentru a practica o anumită activitate sau decizia privind gradul de reliefare a unei caracteristici. Prin intermediul unui test se pot stabili combinaţii individuale ale anumitor trăsături ale personalităţii (*profil de personalitate*). Un alt domeniu de utilizare a testului psihologic este cel al diagnozei longitudinale. În acest caz, se cere stabilirea transformărilor unei caracteristici psihice de-a lungul unui interval de timp (*profil de devenire* sau *profil de evoluţie*), la un anumit subiect sau la grupe de subiecţi. De asemenea, un test poate fi pus şi în serviciul cercetării experimentale (stabilirea de transformări ale unei caracteristici psihice prin varierea sistematică a condiţiilor, la subiecţi separaţi sau la grupe de subiecţi). Sub numele de teste circulă şi o serie de creaţii care satisfac nevoile consumului popular de divertisment (sub formă de problematizări uşoare) şi care au un succes deosebit pe piaţă.

Testele trebuie construite sau selectate dintre cele deja existente, în funcţie de scopul special vizat prin folosirea lor. Chiar în situaţii asemănătoare, utilizarea aceluiaşi test se poate dovedi inadecvată. Utilizarea profesionistă a unui test presupune o bună cunoaştere a condiţiilor în care rezultatele dobândite prin aplicarea lui pot fi relevante. Această cunoaştere este perfecţionată prin acumulări de natură teoretică şi practic-aplicativă. După Paul Kline, reputat specialist în domeniu, construcţia unui test psihologic presupune calităţi care aparţin atât omului de ştiinţă (exact, metodic), cât şi artistului (dezinvolt, fantezist).

Psihologul care îşi propune să construiască un nou test nu este mulţumit de exemplarele existente până în momentul deciziei sale – sau este posibil ca testele existente să nu-i fie accesibile.

De altfel, hotărârea de a utiliza un anumit tip de test trebuie să fie fundamentată şi prin prisma opticii economice. Consultarea cataloagelor principalilor producători de probe psihologice este o etapă obligatorie pentru profesionistul care trebuie să utilizeze un anumit tip de test pentru un anumit scop, deoarece, astfel, acesta poate găsi *„gata făcut"* un test care să corespundă aşteptărilor sale. Nu are rost să facem ceea ce s-a făcut deja. Dar există şi posibilitatea ca un psiholog să nu-şi poată permite, din motive financiare, achiziţionarea unui test. Tot o optică economică de analiză este estimarea costurilor personale ale construirii unui test. Din cunoaşterea modului în care au fost construite o serie de teste foarte cunoscute, putem aprecia că totalul cheltuielilor materiale şi temporale este apreciabil, atât pentru varianta iniţială, cât şi pentru eventualele perfecţionări şi adaptări, deoarece utilizatorul unui test trebuie să aibă grijă permanent să actualizeze criteriile de valabilitate. Neglijarea unor asemenea exigenţe este semnul lipsei de profesionalism (importanţa informaţiei diagnostice face să se vorbească despre o nouă elită, elita persoanelor specializate în construcţia testelor). Amintim faptul că lui Lipot Szondi i-au trebuit 23 de ani ca să definitiveze proba proiectivă care îi poartă numele. Timp considerabil a fost alocat de autori şi pentru realizarea unor probe ca *Stanford-Binet*, *WAIS* şi *WISC*, *16 PF*, *MMPI*, *CPI* etc. Perfecţionările acestor probe se întind pe o perioadă de timp de ordinul deceniilor. De exemplu, Stanford-Binet a fost perfecţionat în 1916, 1937 şi 1961 (potrivit informaţiilor pe care le avem în această privinţă). Autori de teste psihologice foarte cunoscuţi au lucrat doar la o singură probă. De exemplu, David Wechsler, Hermann Rorschach, Max Lüscher, Alfred Binet şi Théodore Simon, Lipot Szondi ş.a. au materializat un singur proiect de construcţie. Toate aceste instrumente psihometrice sunt creaţii de anvergură în istoria psihometriei.

O parte dintre marile corporaţii, care întreţin servicii psihologice cu personal de specialitate bine pregătit, îşi construiesc propriile probe psihologice pentru uz exclusiv (după cum afirmă Paul Horst). Este bine de ştiut că popularitatea unui test nu este, în mod necesar, indiciu al calităţii sale. În multe cazuri, apariţia concurenţei între firmele creatoare de teste determină o reclamă agresivă, care exagerează calităţile testelor.

Un psiholog care-şi propune un scop mai modest, bine precizat, de a construi un test cu aplicabilitate particularizată, uşor de realizat, poate aloca un timp mult mai redus pentru finalizarea unei construcţii psihometrice. Chiar dacă achiziţionăm un test psihologic, nu suntem scutiţi de griji, mai ales în România, deoarece o atenţie deosebită trebuie acordată adaptării acestuia, dovedindu-se inadecvată preluarea unor sisteme de norme naţionale existente în ţara producătoare. Studiul colecţiei de lucrări ştiinţifice cuprinse în lucrarea *Mental test and cultural adaptation* (1972), editată de L.J. Cronbach şi P.J. Denth, oferă un bun prilej de a inventaria factorii care pot influenţa rezultatele testelor, din moment ce acestea sunt aplicate în medii culturale diferite.

Profesionistul interesat de construirea unui instrument psihometric trebuie să aibă serioase disponibilităţi în materie de practică experimentală, deoarece lansarea unei probe psihologice presupune examinarea prealabilă a unui mare număr de subiecţi şi prelucrarea corespunzătoare a datelor. În ţările cu puternice asociaţii profesionale, exigenţele faţă de calităţile unui test psihologic sunt deosebit de mari. Este cazul American Psychological Association (APA) care posedă norme severe de acceptare a produselor metodologice, reevaluate periodic.

Un psiholog care doreşte să construiască un test psihologic, în condiţiile actuale, trebuie să mai ştie că activitatea sa va fi una de colaborare cu alţi profesionişti. Vremea cercetătorului solitar, izolat în liniştea cabinetului sau laboratorului său, studiind şi creând

produse ştiinţifice de excepţie, a trecut. În lumea în care trăim, schimburile informaţionale sunt un fapt curent, greu de evitat pentru realizarea unor creaţii ştiinţifice valide. Accesul la mari surse de date este facilitat de internet şi de alte reţele informaţionale. Produsele *software* de tip *bază de date* sunt numeroase, fără a fi prohibite financiar. Dată fiind această situaţie tehnologic-informaţională favorabilă, psihologul poate lua legătura foarte rapid cu alţi colegi care au preocupări asemănătoare, din orice parte a lumii. În acest fel putem evita producerea unui instrument existent deja. Recurgerea la mijloacele pe care ni le pune la dispoziţie informatica modernă este obligatorie pentru a deveni eficienţi.

Lucrurile se schimbă total în cazul în care psihologul activează în cadrul unei edituri având ca obiect principal de lucru producerea şi comercializarea testelor psihologice. În acest caz, cunoaşterea procedurilor de construcţie a testelor are o importanţă aparte, produsele create fiind răspândite unui mare număr de beneficiari. Responsabilitatea activităţii profesionale creşte foarte mult. În România, nu există încă edituri specializate în producerea şi comercializarea testelor, dar lucrurile vor evolua, cu siguranţă, şi în această privinţă.

Un procedeu mai practic de a lansa un nou test psihologic este acela de a modifica semnificativ sau de a recombina creaţiile deja existente. În acest caz, este important gradul de modificare şi trebuie să se respecte dreptul de autor, care operează juridic cu deosebită severitate. David Wechsler a practicat acest procedeu cu un succes remarcabil, dar a avut acceptul autorilor testelor pe care le-a combinat într-un mod foarte inteligent.

2. Etape în construcţia unui test

Construcţia unui test psihologic cunoaşte o serie de etape, care diferă neesenţial de la un autor la altul. Dany Laveault şi Jacques Grégoire (1997, pp. 79-85) disting, astfel, cinci etape:

- determinarea utilizărilor prevăzute de test;
- definirea a ceea ce se doreşte a măsura;
- crearea itemilor;
- evaluarea itemilor;
- determinarea proprietăţilor metrice ale testului definit.

Robert Gregory (1996, p. 129) stabileşte următoarele etape:

- definirea testului;
- stabilirea unei metode de scalare;
- construcţia itemilor;
- testarea itemilor;
- revizia testului;
- publicarea testului.

A.F. Vukovici (1964) consideră că, în cazul construcţiei testului, trebuie realizate o serie de prescripţii de procedură referitoare la:

- populaţia examinaţilor;
- materialul testului care trebuie să stimuleze examinatul spre un comportament specific;
- stabilirea indicaţiilor testului prin care este expusă problema cercetată sau prin care subiectul este invitat la o anumită interacţiune cu materialul testului;

- stabilirea diverselor condiţii în care sunt oferite instructajul şi materialul testului ;
- delimitarea comportamentelor considerate relevante în plan diagnostic ;
- stabilirea unor reguli de ordonare şi scheme de clasificare, diferitele moduri de comportament semnificative fiind împărţite în categorii ;
- calcularea frecvenţelor claselor de valori brute ;
- sistemul de norme şi referinţe, servind interpretării şi comparării valorilor brute.

Paul Kline indică două metode de abordare a construcţiei testelor :
- metoda bazată pe analiza itemilor sau pe analiza factorială ;
- demersul empiric care a fost pus la punct de creatorii testului clinic MMPI.

În primul caz, se folosesc intens analize statistice având drept scop crearea unui test unidimensional, care îşi propune să măsoare o dimensiune sau un factor. După ce ansamblul de itemi a fost pus la punct, este necesar să se demonstreze validitatea construcţiei. Această primă manieră de abordare a construcţiei testului are următoarele subetape : stabilirea numărului de itemi ; stabilirea eşantioanelor de încercare ; administrarea testului ; analiza itemilor ; selecţia itemilor după o primă analiză a itemilor ; rescrierea şi validarea itemilor.

A doua manieră de abordare presupune depăşirea a două probleme cruciale : compoziţia grupului-criteriu şi stabilirea semnificaţiei scalelor construite plecând de la aceste criterii.

J.J. Bernier şi B. Pietrulewicz (1997, p. 235) stabilesc opt etape pe care le considerăm un demers clar :

1. Specificarea scopului.
2. Exprimarea acestui scop în termeni discursivi.
3. Precizarea conţinuturilor, cu subetapele : a) descrierea conţinuturilor şi a abilităţilor ; b) definirea unei trăsături şi c) analiza activităţilor şi definirea componentelor trăsăturii şi a criteriilor.
4. Redactarea itemilor.
5. Verificarea itemilor şi analiza pretestului (discriminarea şi dificultatea, pertinenţa altor posibilităţi).
6. Compunerea formei finale a testului.
7. Standardizarea (administrare, limite de timp, moduri de corecţie).
8. Analiza tehnică a testului (precizie, validitate, norme).

Prezentăm în continuare, sub formă schematică, etapele construcţiei unui test, în viziunea lui Helmuth Benesch (1995, p. 354) :

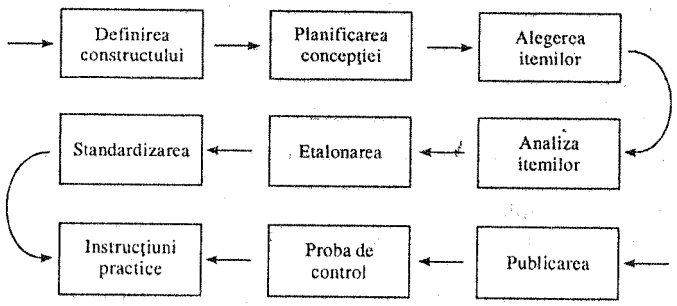
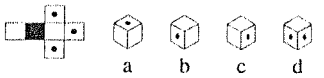


Figura 13. *Etapele construcţiei testului* (apud *Helmuth Benesch, 1995*)

După această trecere în revistă a unor etape în construcţia testului, ne permitem să facem câteva remarce :

- Nu toţi autorii stabilesc o etapizare a construcţiei testului (de exemplu, Lienert, Guthke, Magnusson, dar aceasta este implicită). Etapele, o dată fixate, au ceva din forma rigidă a unei construcţii scolastice.
- Construcţia unui test nu este niciodată încheiată. Elaborările reputate din istoria testului psihologic au suferit repetate revizii, începând cu scalele Binet-Simon şi Terman-Merrill. McKinley şi Hathaway (*apud* Spreen, 1963) atrag atenţia, în manualul de prezentare MMPI, că proba lor nu este terminată şi că gruparea în scale poate continua după alte criterii. Probele lui David Wechsler au avut o elaborare de durată şi multiple revizuiri. La fel s-a întâmplat cu *16 PF* al lui R.B. Cattell. Matricele progresive Raven au cunoscut îmbunătăţiri în scopul onorării destinaţiei psihodiagnostice.
- Anumite etape ale construcţiei testului devin preocupări autonome, de exemplu analiza de itemi (şi, chiar în cadrul acesteia, analiza distractorilor). Într-o lucrare de referinţă din domeniul psihometriei (Dickes *et al.*, 1994, p. 89) se afirmă că o teorie a distractorilor urmează să fie construită.

Pentru exemplificarea unui item dintr-un test privind relaţiile spaţiale, prezentăm următoarea situaţie problematică :



în care i se cere unui subiect să indice care dintre cele patru figuri din dreapta se poate obţine prin plierea cubului desfăşurat din stânga.

Un exemplu de itemi tipici pentru matricea progresivă Raven este dat în figura următoare :

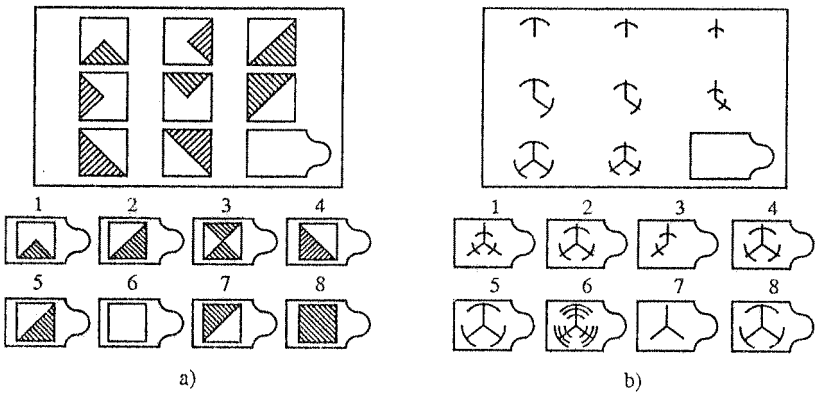


Figura 14. *Forme de stimulare a comportamentului de rezolvare pentru itemii matricei progresive Raven*

Dintre aspectele care condiţionează construcţia unui test bun, ne vom opri la concepţia aflată la baza construcţiei testului, exigenţele pentru crearea unui bun item, asigurarea criteriilor de valabilitate, tipurile de sarcină şi dispozitivele de răspuns, precum şi la eşantioanele pe care se verifică testul.

3. Teoria aflată la baza construcţiei testului psihologic

La baza construcţiei oricărui test psihologic se află o teorie a personalităţii sau o teorie privind natura caracteristicii investigate. Referindu-se la principiile metodologice după care s-ar ordona metodele de evaluare în psihologie, Gordon Allport (1991, p. 395) consideră că în prim-plan se află „teoriile care stau la baza unui instrument (dacă derivă din psihanaliză, asociaţionism, analiză factorială, psihologia genetică, behaviorism)".

Teoria aflată la baza construcţiei unui test este implicită sau explicită. În unele cazuri, teoria are componente provenind din mai multe curente psihologice. Zena Helman, profesor de psihopatologie la Université Lille-III spune, în această privinţă: „*Psycho-diagnostik*-ul pe care Rorschach îl publică în anul 1910 se găseşte la o răspântie de drumuri. Acest volum dens şi foarte bogat, dar în mare parte empiric – după cum o recunoaşte chiar autorul – poartă pecetea mai multor curente: psihologia asociaţionistă moştenită din secolul al XIX-lea, psihanaliza care apăruse prin opera lui Freud de circa un sfert de secol şi psihopatologia structurală care îşi are începutul în acea epocă" (1995-1996, p. 113).

Într-un număr de cazuri, aspectelor teoretice li se consacră un spaţiu foarte restrâns, dar în alte cazuri spaţiul teoriei cuprinde lucrări în mai multe volume. Ne referim la baza teoretică a testelor proiective Rorschach, Lüscher, Szondi. *Der 4 Farben Mensch oder Weg zum Gleichgewicht* a lui Max Lüscher cuprinde în sute de pagini un lung excurs despre specificul uman şi echilibrul său intern, psihologia şi filosofia celor patru culori, normele etice fundamentale, tacticile de apărare contra agresivităţii cromatice.

Manualul testului Rorschach, prezentat de către M. Loosli-Usteri, conţine proceduri asemănătoare (de dimensiuni mai reduse) despre studiul teoretic al interpretărilor referitoare la culoare, studiul teoretic al interpretărilor referitoare la clar-obscur, analiza interpretării formelor umane şi animale etc.

Testul proiectiv creat de Szondi (una dintre cele mai enigmatice figuri din istoria psihodiagnosticului) impresionează prin discrepanţa dintre teoria genetică a alelelor recesive, depăşită actualmente de evoluţia ştiinţei biologice, şi puzderia de formule pulsionale rezultând din aceasta (*cf.* Legrand, 1979).

În general, o teorie este un corpus sistematizat de aserţiuni în legătură cu un domeniu al realităţii, care rămâne deschis confirmării sau punerii în discuţie. O întreagă sistematică operează în studiul teoriilor ştiinţifice. Karl Popper susţine că orice teorie experimentală este falsificabilă, adică poate fi infirmată de evoluţia investigaţiilor ştiinţifice. Ca atare, teoria ştiinţifică nu are statutul unui adevăr ştiinţific definitiv. Prin celebra sa teorie a paradigmelor, Thomas Kuhn precizează foarte clar durata de viaţă a teoriilor ştiinţifice în cadrul comunităţii specialiştilor dintr-un domeniu. Folosirea de durată a termenilor unei teorii poate dăuna evoluţiei ştiinţifice prin subordonarea metodei de cercetare poziţiei teoretice. Françoise Parot şi Marc Richelle (1995, p. 195) se exprimă foarte plastic în această privinţă: „Când omul de ştiinţă îşi proclamă supunerea în faţa faptelor, el nu poate ignora că le priveşte, le înţelege, le caută prin formulări teoretice ca printr-o lentilă uşor colorată care, chiar prin proprietăţile sale, decupează şi colorează realitatea într-un anumit mod". Despre implicaţiile categorizării s-au scris lucrări care constituie subiect de interes pentru specialiştii în metateorii.

O teorie poate fi menţinută artificial, prin operarea cu anumite trucuri metodologice, dintre care selectarea doar a faptelor care o confirmă, întreruperea cercetării în momentul în care rezultatele sunt favorabile teoriei etc.

Aşa cum am menţionat şi în secţiunea despre validare, investigarea prin intermediul testului psihologic nu permite examinarea unor concepte „pure", ci doar a unor modele ale acestora (constructe).

O excelentă prezentare a modelelor psihologice de descriere a activităţii aparţine unui grup de autori francezi, format din Claude Bonnet, Christiane Kekenbosch, Anh Nguyen Xuan, J.F. Richard şi Charles Tijus[1]. În concepţia lor, modelele în discuţie au cinci caracteristici majore:

1. Un model este o particularizare a unei teorii de funcţionare cognitivă. El se ataşează unei teorii, dar se aplică unui tip de situaţie particulară şi face ipoteze suplimentare în raport cu teoria a cărei prelungire este.
2. Un model este o formalizare şi, ca atare, se exprimă într-un limbaj formalizat. Acest gen de limbaj este fără ambiguitate şi permite calcule, adică deducţii automate. O bună perioadă de timp, limbajul logic şi matematic erau singurele care permiteau calcule. Însă informatica a permis dezvoltarea altor limbaje formalizate care fac posibile, pe lângă calcule numerice, şi calcule simbolice, adică deducţii asupra cunoştinţelor şi a semnificaţiilor.
3. Un model este schematic, altfel spus reprezintă o simplificare a realităţii care nu reţine decât aspectele esenţiale (acestea sunt aspectele care, în principiu, sunt suficiente pentru a prezice un comportament). Caracterul schematic este şi o caracteristică a teoriilor, deci atât modelele, cât şi teoriile revendică ideea potrivit căreia aceste simplificări conservă trăsăturile esenţiale ale realităţii în raport cu tipul de comportament pe care îl explică.
4. Un model furnizează predicţii validabile prin date experimentale, deci el trebuie să „reziste" la o serie de provocări ale realităţii. Un model psihologic conţine o descriere a situaţiei la care se aplică el (şi care nu reţine decât un anumit număr de elemente pertinente). Modelele conţin o descriere a mecanismelor de tratament informaţional pentru acele tipuri de situaţii cu ajutorul cărora se poate prezice comportamentul posibil al unui subiect, care reacţionează după model. Un astfel de procedeu de tratament informaţional se numeşte *simularea comportamentului*.
5. Investigarea, prin intermediul unui model, se realizează în proporţii variabile. În multe cazuri, atât modelele, cât şi teoriile sunt implicite, iar autorul testului nu insistă prea mult asupra părţii de fundamentare teoretică a instrumentului său. Tipul de solicitare pe care îl presupune sarcina psihologică trebuie să fie exprimată în forme concrete răspunzând exigenţei teoretice.

1. Studiul se intitulează „Mesure et modélisation" şi face parte din volumul *Cours de Psychologie* (coord. R. Ghiglione şi J.F. Richard), Dunod, Paris, 1995, pp. 268-384.

Practic, orice test aflat în construcţie trebuie să verifice o ipoteză. Pentru a supune o ipoteză la proba faptelor, este necesar să traducem noţiunile prin operaţii de măsură sau observaţii care le vor da un conţinut empiric. Dacă o experienţă implică viteza de tratare a informaţiei, trebuie aleasă o operaţie care să permită măsurarea respectivei viteze. Există însă mai multe operaţii, deja utilizate de diferiţi cercetători, pentru a efectua o măsurare. Aceste procedee diferite furnizează măsuri în corelaţii relativ slabe. Mijlocul de a defini cel mai bun procedeu nu este evident. Pentru a avea certitudine, trebuie comparate toate procedeele cu o măsură „adevărată" a vitezei de tratare a informaţiilor (dacă se cunoaşte o măsură „adevărată", nu mai trebuie pusă problema privind cel mai bun procedeu). În fapt, alegerea unei operaţii se face în fiecare cercetare sub efectul ipotezelor, care sunt rar explicite. Se poate ajunge în situaţia în care alegerea procedeelor diferite pentru a operaţionaliza aceeaşi noţiune să aducă diferenţe în rezultatele obţinute de cercetători diferiţi.

Maurice Reuchlin (1992, pp. 33-34) oferă exemple foarte interesante privind realizări diferite ale unei aceleiaşi noţiuni. Realizările aparţin mai multor cercetători care au studiat sarcinile în al căror rezultat intervine viteza cu care subiectul tratează informaţia furnizată. Procedeele poartă numele autorilor lor.

- **Paradigma lui T. Nethlebeck** (timpul de inspecţie) : două linii verticale de lungimi puţin diferite sunt prezentate subiectului într-un timp foarte scurt, iar acesta trebuie să semnaleze dacă cea mai scurtă se situează în dreapta sau în stânga. Se determină perioada de timp necesară subiectului pentru a furniza un procentaj de răspunsuri exacte (de exemplu, 95%).
- **Paradigma lui W. Hick** (timpul de decizie) : subiectul trebuie să stingă lumina chiar în momentul în care ea se aprinde, apăsând un buton. Becul care se aprinde poate fi singura sursă de lumină prezentată sau aceasta trebuie aleasă la întâmplare de către experimentator (ori de un program ce pilotează experienţa) într-o serie de 2, 4 sau 8 surse susceptibile de a se aprinde. Incertitudinea subiectului creşte o dată cu creşterea numărului de surse luminoase. Timpul de reacţie creşte liniar în funcţie de această măsură de incertitudine. Panta dreptei care reprezintă creşterea respectivă este un parametru reţinut.
- **Paradigma lui S. Sternberg** (viteza de acces a memoriei pe termen scurt) : subiectul observă timp de mai multe secunde o serie de cifre de la 2 la 7. Apoi seria dispare şi este prezentată o cifră. Subiectul trebuie să semnaleze cât mai rapid posibil, apăsând pe unul dintre cele două butoane aflate la dispoziţia sa, dacă cifra figurează sau nu în seria iniţială. Parametrii reţinuţi sunt ordonata la origine şi panta dreptei reprezentând creşterea timpului în funcţie de numărul de cifre din serie.
- **Paradigma lui M. Posner** (viteza de acces la memoria de lungă durată). Subiectului i se prezintă perechi de litere precum AA, Aa, Ab. Este măsurat mai întâi timpul de reacţie necesar subiectului pentru a semnala, după prezentarea fiecărei perechi, dacă cele două litere sunt identic capitalizate (ambele litere sunt majuscule sau minuscule). Timpul de reacţie este mai mare când, ulterior, i se cere subiectului să semnaleze dacă cele două litere sunt semantic diferite sau nu. Diferenţa dintre cei doi timpi este luată ca normă a timpului de acces la memoria de lungă durată care intervine pentru a constata dacă „A" şi „a" sunt două litere semantic identice, nefiind nevoie să intervină în constatarea unei identităţi fizice (cum ar fi într-un cuplu AA, de exemplu).

Viteza de tratare a informaţiei pare să intervină în rezultatul celor patru probe. Dar ea intervine în concurenţă cu alţi factori variind de la o sarcină la alta. Autorul poate să continue să definească în cercetarea sa viteza de tratare a informaţiei printr-una dintre aceste paradigme.

4. Formularea unui item bun

În privinţa formulării unui item bun, considerăm necesar să amintim că nu orice test se compune din itemi, în unele cazuri aflându-ne în situaţia unui continuum solicitativ (cum este cazul probelor pentru examinarea asocierii şi disocierii manuale - DM, sinusoida Bonnardel). Itemul, în calitate de componentă informaţională bazală, se întâlneşte la chestionarele de personalitate, la testele de cunoştinţe şcolare, testele verbale şi nonverbale de inteligenţă. Exigenţele în privinţa formulării itemilor se află în continuă creştere, tot aşa cum sunt şi eforturile de cercetare privind selectarea unui bun item, specificul informaţional al fiecărui tip de item, determinarea numărului optim de itemi care compun un test. Transferul preocupărilor ştiinţifice către aceşti „atomi" ai testului se pare că este o mutaţie deosebit de benefică pentru construirea unui instrument de măsură adecvat domeniului psihodiagnosticului. Această analiză a testului, pornind de la şiruri de informaţii elementare, aduce un spor informaţional considerabil şi de o mare diversitate. Situaţia pare să exemplifice un pasaj din capitolul final, intitulat „Integronul", al cărţii *Logica viului*, scrisă de François Jacob, laureat al Premiului Nobel pentru Medicină : „Varietatea lumii vii, extraordinara diversitate de forme, de structuri şi de proprietăţi observate la nivel microscopic, se întemeiază pe o combinatorică a câtorva specii moleculare, adică pe o simplitate a mijloacelor la nivel microscopic" (1972, p. 323).

Pentru chestionarele de personalitate, se recomandă (vezi Kline, 1995, pp. 122-123) respectarea unor exigenţe formale în formularea itemilor, pe lângă cele de conţinut :

- evitarea folosirii unor termeni care sunt, în mod vădit, dezirabili sau indezirabili din punct de vedere social. Mai ales în cadrul unor examene de evaluare, subiecţii vor da acele răspunsuri pe care le consideră ca fiind apreciate de către examinator ;
- se impune ca la o parte dintre itemi să se răspundă cu „da" şi la altă parte cu „nu" (răspunsul valorizat de chestionar), pentru a se evita aşa-numitul „efect de aprobare" ;
- se impune, de asemenea, redactarea cât mai simplă posibil a itemilor care trimit mai curând la comportamente, decât la sentimente sau la emoţii ;
- evitarea referinţelor la termeni, precum : „frecvent", „de obicei", „adesea" etc. ;
- fiecare item nu trebuie să conţină decât un element de apreciat (nu se pot admite formulări de genul „Îmi plac la fel de mult literatura şi istoria").

O parte dintre autori nu recomandă răspunsuri intermediare, între „da" şi „nu", deoarece aceasta ar atrage indecizia. Considerăm că doar la unele întrebări dintr-un chestionar de personalitate se poate răspunde prin „da" sau „nu", „adevărat" sau „fals". Se poate afirma, de asemenea, că nu există o formă perfectă de exprimare la item care să satisfacă absolut toţi constructorii de teste.

Nu se poate susţine nici existenţa unei mărimi standard a testului, legată de numărul de itemi. Dacă un criteriu de economicitate este presant, va trebui redus numărul de

itemi, dar nu dincolo de limita unor valori minimale asiguratorii pentru criteriile de apreciere a testului. Testele care au ca exigenţă o mare fidelitate necesită un număr mai mare de itemi.

Din literatura de specialitate parcursă, doar Paul Kline face referire la numărul minim de itemi pe care trebuie să-i conţină un test, şi anume 10. Considerăm cifra exagerat de mică, gândindu-ne la foarte redusa putere de discriminare a unei probe cu o amplitudine mică a punctajului. În faza preliminară de construcţie a testului, majoritatea autorilor de specialitate recomandă ca numărul de itemi să fie mărit cu 15-30% faţă de numărul propus pentru varianta finală.

Partea psihometrică de construcţie a unui item va avea în vedere stabilirea unor indicatori şi a unor corelaţii: indice de dificultate, indice de segregare, de validitate, de fidelitate şi corelaţia itemului cu rezultatul total prin intermediul analizei punct-biseriale (ilustrată deja de noi).

Cel care concepe un test psihologic este obligat să procedeze în funcţie de accentul pus pe validitatea de conţinut sau validitatea predictivă.

5. Asigurarea criteriilor de valabilitate a unui instrument psihometric

În activitatea de construire a unui test psihologic, există câteva proceduri esenţiale. Ne referim în mod deosebit la cele de asigurare a obiectivităţii, fidelităţii şi validităţii. Este foarte dificil, dacă nu chiar imposibil, să se indice rigid paşii ce trebuie în mod necesar parcurşi pentru asigurarea acestor criterii de valabilitate a testului. Dacă la stabilirea obiectivităţii sau a fidelităţii lucrurile par mai clare, în cazul validităţii intervine un mare procent de relativitate. Pentru a putea asigura criteriile de valabilitate a testului, trebuie să analizăm foarte limpede scopul pe care ni-l propunem în construcţia unui nou test. În funcţie de acest scop, putem particulariza calea pe care trebuie să o urmăm. Pentru asigurarea criteriilor de apreciere a unui test, pot fi aplicate diferite demersuri pertinente. Un al doilea lucru pe care trebuie să-l avem în vedere este că aceste criterii de apreciere nu pot fi stabilite o dată pentru totdeauna, activitatea constructorului trebuind să fie caracterizată de o anumită continuitate. Testul, asemenea unor produse tehnologice moderne, poate să dobândească un grad de uzură morală, cu trecerea anilor, adică să nu mai satisfacă cerinţele pentru care a fost creat, de multe ori nemaifiind necesar să fie utilizat, şi asta deoarece nu mai este urmărită calitatea diagnosticată care este depăşită de evoluţia ideilor şi a realităţilor practice. Ca atare, intensitatea efortului nostru trebuie să fie pe măsură, asigurându-se integrarea lui în contextul pragmatic şi ştiinţific.

Pentru demersul de asigurare a fidelităţii unui test psihogic, am considerat potrivită exemplificarea dintr-o reputată lucrare de prezentare a testelor (vezi Brickenkamp, 1997, pp. 117-118). Se prezintă rezultatele obţinute pe diferite componente ale testului, precum şi o serie de deficienţe (prezentarea acestora poate fi considerată un act deontologic).

Fidelitatea. HAWIE-R (Hamburg-Wechsler-Inteligenztest für Erwachsene. Revăzut în 1991). Editor: U. Tewes.

Analiza de itemi. În manualele de utilizare sunt enumerate informaţii (date) despre media şi discriminarea (0,08 până la 0,66) tuturor itemilor.

Consistenţă internă

Partea verbală

- r_{tt} = 0,57 până la 0,92 pentru nouă grupe de vârstă (16 până la 74 de ani);
- r_{tt} = 0,60 până la 0,95 pentru toate grupele de vârstă şi partea verbală în general.

Partea de randament

- r_{tt} = 0,60 până la 0,95 pentru toate grupele de vârstă şi pentru cinci subteste;
- r_{tt} = 0,89 până la 0,97 pentru toate grupele de vârstă şi pentru partea de randament, în general.

Ansamblul testului

- r_{tt} = 0,95 până la 0,97 pentru toate grupele de vârstă şi testul total.

Cu 0,95 subtestul ZS dovedeşte cea mai înaltă consistenţă medie, în timp ce subtestul BE, cu 0,72, cele mai scăzute valori.

Stabilitate: nu există informaţii

Eroarea standard de măsură

Partea verbală

- s_e = 0,87 până la 1,64 (unităţi de punctaj) pentru toate cele nouă grupe de vârstă şi şase subteste;
- s_e = 2,60 până la 3,97 (unităţi IQ) pentru toate grupele de vârstă şi pentru partea verbală în ansamblu.

Partea de randament

- s_e = 0,64 până la 1,92 (unităţi de punctaj) pentru toate grupele de vârstă şi pentru cinci subteste
- s_e = 3,97 până la 4,74 (unităţi IQ) pentru toate grupele de vârstă şi pentru partea de randament în ansamblu.

Partea totală

- s_e = 2,60 până la 3,35 (unităţi IQ) pentru toate grupele de vârstă şi pentru testul în ansamblu.

Omogenitate

Intercorelaţiile subteste în partea verbală: r_{tt}=0,10 până la 0,82; în partea de randament: r_{tt}=0,05 până la 0,68; între partea verbală şi partea de randament: r_{tt} = 0,01 până la 0,70.

Pe baza intercorelaţiilor parţiale la subteste semnificative, ar trebui să se enunţe o interpretare de profil.

Pentru validitate, vom prezenta în continuare două demersuri de asigurare a acesteia: un demers pragmatic simplu şi un demers pragmatic complex. Prin intermediul lor, vom întări ideea potrivit căreia stabilirea validităţii unor intrumente psihometrice se poate face la diferite niveluri de complexitate, fără afectarea calităţii.

Un *demers pragmatic simplu* pentru stabilirea validităţii concurente considerăm că este cel realizat de Marioara Zâpârţan într-o lucrare (1990) consacrată cunoaşterii factorilor de personalitate în orientarea şcolară şi profesională a elevului. Autoarea a construit o scală de atitudini faţă de meseria de vânzător, deci un instrument de investigaţie foarte strict delimitat în privinţa scopului psihodiagnostic. Faptul că ne aflăm în prezenţa unui singur instrument psihodiagnostic, că se urmăreşte asigurarea unui singur fel de validitate şi că procedeele statistice ţin preponderent de statistica bivariată ne determină să considerăm procedura de validare ca fiind un *demers pragmatic simplu*. Sintagma nu este încărcată de semnificaţie peiorativă. Autoarea utilizează date provenind de la un eşantion de 256 de subiecţi. Pentru instrumentul de investigaţie psihodiagnostică amintit, Marioara Zăpârţan consideră că cea mai adaptată formă de validitate este cea concurentă, ca o primă formă de validare necesară. În funcţie de contextul în care a lucrat, autoarea apreciază că tipul de criteriu cel mai potrivit este criteriul subiectiv. La această concluzie s-a ajuns după ce s-au analizat posibilităţile de utilizare a unui criteriu obiectiv. Autoarea constată că interesul pentru meseria de vânzător, exprimat verbal şi atitudinal, prin orientarea spre această profesiune, este întâlnit într-o proporţie ridicată la elevii cu un randament şcolar mediu sau sub nivelul respectiv. Tot la această categorie de subiecţi este prezentă adeziunea afectiv-motivaţională pentru profesiunea de vânzător. De aici doza de reţinere în utilizarea exclusivă a randamentului şcolar drept criteriu obiectiv.

În scopul estimării (şi diminuării) incertitudinii unor aprecieri, s-a apelat la un grup de evaluatori cu niveluri diferite de calificare (profesorul, dirigintele, instructorul de practică, şefii de magazin, câţiva vânzători sau colegi).

S-a utilizat modelul criteriului compozit, compus din subcriterii, precizându-se dimensiunile componente ale acestuia. Subcriteriile au fost stabilite pornindu-se de la studiul meseriei de vânzător, care cuprinde analiza muncii şi trăsăturile psihice necesare în profesiunea respectivă. S-au fixat patru criterii parţiale (aptitudini şi pregătire specifice meseriei, interes verbalizat şi manifestat, calitatea lucrărilor şi disciplina). Înainte de efectuarea aprecierilor, s-au purtat discuţii cu evaluatorii (pornind de la studiul profesiunii de vânzător) în legătură cu conţinutul fiecărui subcriteriu, în scopul de a se ajunge la o justă înţelegere a semnificaţiei fiecărei componente din criteriul compozit. În acest scop, evaluatorilor li s-au oferit precizări de apreciere a subiecţilor la fiecare componentă luată separat, în funcţie de gradul în care caracteristica în discuţie este mai mult sau mai puţin prezentă la vânzători. Criteriul compozit cuantificat constituie o sumă ponderată a cotelor subcriteriilor (s-au acordat ponderi egale criteriilor parţiale şi s-a făcut o medie a unor astfel de scoruri parţiale). Criteriul compozit a fost analizat sub aspectul omogenităţii sale prin intermediul matricei intercorelaţiilor dintre criteriile parţiale. S-a trecut apoi la verificarea fidelităţii criteriului prin analiza de varianţă; cu cât diferenţele interindividuale sunt mai mari, iar diferenţele intraindividuale mai mici, cu atât fidelitatea măsurătorii este mai mare. Deoarece grupele prea omogene de subiecţi distorsionează varianţa, în lotul de cercetare a fidelităţii criteriului au fost incluse *grupe extreme* formate din subiecţi ce aderă la meseria de vânzător (subiecţi cu atitudini pozitive faţă de profesiunea în cauză) şi subiecţi refractari (atitudine negativă faţă de această profesiune). Subiecţii din aceste grupuri, în număr de 40, au fost supuşi evaluării de şase ori de-a lungul unei jumătăţi de an. Analiza de varianţă a fost realizată cu modelul descris de J.M. Faverge.

Autoarea încearcă să elimine posibilele distorsionări ale criteriului, evitând apariţia unor erori sistematice legate de caracteristicile grupului, de complezenţa evaluatorilor şi

de contaminarea datorată cunoaşterii predictorului. S-a căutat ca eşantionul de validare să-şi păstreze omogenitatea sub raportul vârstei, al nivelului de şcolarizare, al menţinerii unui echilibru între numărul de elevi proveniţi din mediul urban şi cei proveniţi din mediul rural. De asemenea, au fost fixate o serie de criterii de alegere a evaluatorilor, în vederea evitării distorsionării rezultatelor prin nerelevanţa calităţii de evaluator. Pentru a nu se contamina valoarea aprecierii prin cunoaşterea rezultatelor de către evaluatori, s-a evitat, în faza exercitării evaluării, să se aducă la cunoştinţa acestor experţi rezultatele la scala de atitudine. Coeficientul de corelaţie, în studiul de validare, a fost completat cu ecuaţia de regresie prin care s-a ridicat gradul de paralelism dintre cele două variabile. Problema predicţiei implică trei variabile: variabila diagnostică sau examenul, care cuprinde datele culese în timpul examinării (x); variabila-criteriu (y), care se referă la ceea ce vrem să prognosticăm pornind de la diagnostic; variabila-pronostic y' care, în funcţie de diagnosticul x, este estimarea pe care psihologul o face cu privire la valoarea necunoscută a lui y. În situaţia exemplificată, există variabilele x şi y.

S-a efectuat o validare predictivă parţială, corelându-se datele obţinute la scara de atitudini cu evaluarea subiecţilor de către experţi la un interval de un an de la data testării. Formele concurente şi predictive de validare au fost întregite cu validarea de conţinut. S-a urmărit desprinderea din totalul itemilor incluşi în construcţia scalei a celor care, prin conţinutul lor, converg spre sarcini vizate de test în ansamblu. Validarea de conţinut ia forma analizei de itemi, în cadrul căreia se realizează analiza proprietăţilor statistice ale itemilor.

Determinarea indicilor statistici (în analiza itemilor) cuprinde:

- gradul de dificultate a itemilor sau *indicele de popularitate*;
- corelaţia item – test;
- valoarea discriminativă a itemilor.

Indicele de popularitate a fost calculat pe baza proporţiei subiecţilor care au răspuns afirmativ sau au aprobat atitudinea exprimată de fiecare item.

Corelaţiile item – test au fost făcute în scopul diferenţierii itemilor discriminativi de cei nediscriminativi. Acestea au fost dublate de corelaţia item – criteriu, pentru dobândirea unui mai mare grad de certitudine în acţiunea de *purificare* a itemilor.

Pentru o diferenţiere optimă a *valorii discriminative a itemilor*, s-a completat corelaţia item – test, efectuată pentru fiecare item în parte, cu corelaţia item – test şi item – criteriu pe categorii de atitudini.

Un exemplu pentru un *demers pragmatic complex* în asigurarea validităţii considerăm că poate fi găsit în cercetarea realizată de Horia Pitariu (1983, pp. 258-262) pentru o baterie de teste destinată selecţiei operatoarelor şi perforatoarelor de cartele. În urma analizei psihologice a profesiei de operator-perforator de cartele, au fost selectate iniţial un număr de zece probe psihologice, dintre care au fost reţinute ca semnificative un număr de cinci: *testul de atenţie susţinută* (AS) creat de Ricossay; *testul de atenţie distributivă* (AD), *testul de abilităţi aritmetice simple* (AA), *testul Domino* (D-48) şi *Tapping* (T). Analiza factorială a testelor a evidenţiat trei factori. Primul factor se referă la capacitatea de efort, de concentrare a atenţiei în condiţii de activitate intensă (factorul *acoperă* 37,65% din varianţă). Testele reprezentate în *factorul I* sunt AS (ritm impus rapid), AD şi AA. *Factorul II* se poate defini ca un factor de structurare perceptuală, opus celui de dexteritate digitală, şi vizează testele AS (al doilea subtest) şi testul D-48 (se *acoperă* 23,38% din varianţă). *Factorul III* este unul de motricitate şi

se referă la testele AS (ritmul impus lent şi ritmul impus rapid) şi testul T (factorul explică 18,93% din varianţă).

Validarea s-a desfăşurat în două etape, prima fiind un studiu-pilot realizat pe un lot restrâns (N=20), iar a doua etapă referindu-se nu atât la un studiu de contravalidare, cât mai ales la operaţia de validare propriu-zisă. Al doilea lot are un volum N=50, fiind alcătuit din operatoare profesioniste. În tabelele create în cercetarea amintită este redată structura celor două loturi care au făcut obiectul validării testelor psihologice, în ceea ce priveşte performanţele obţinute de către subiecţi la teste.

Pentru stabilirea validităţii probelor, s-a considerat că se poate obţine un criteriu relevant, care să poată discrimina operatoarele între ele, combinând reuşita profesională cu aprecierile şefilor direcţi. Acest lucru a fost posibil deoarece lotul avea dimensiuni reduse, iar instituţia beneficiară ţinea evidenţa performanţelor zilnice ale fiecărei operatoare. În consecinţă, s-a elaborat o scară de apreciere de la 1 la 10. În funcţie de aceste elemente, s-a fixat un criteriu comun care apărea sub forma unei medii aritmetice a notelor. În urma validării au rezultat coeficienţii de corelaţie, prezentaţi în tabelele din lucrare.

Interpretarea coeficienţilor de validitate sub aspect statistic este completată cu o analiză de conţinut, în scopul prevenirii apariţiei unor surse de eroare (s-a luat în considerare faptul că lotul studiat nu era prea mare). Dacă experimentul-pilot a reliefat o serie de tendinţe şi a confirmat în linii generale ipoteza cu privire la alegerea probelor presupuse a fi predictive asupra eficienţei activităţii de perforare a cartelelor, contravalidarea a vizat realizarea unor corecţii pentru asigurarea unei relevanţe sporite a datelor obţinute. Când s-a lucrat cu lotul de 50 de subiecţi care reprezentau operatoarele profesioniste, s-a ajuns la reconsiderarea probei *Tapping* care, anterior, dăduse rezultate îndoielnice. Testul *D-48* posedă o validitate redusă (din tabelele construite de autor, se observă că acest test corelează foarte slab cu ansamblul celorlalte teste). Totuşi, după operarea corecţiei, s-a observat că acesta avea o pondere semnificativă din punctul de vedere al contribuţiei în cadrul bateriei de teste.

Aprofundarea analizei privind structura bateriei de teste prin studiul intercorelaţiilor relevă o independenţă satisfăcătoare, procentul de varianţă acoperit de teste luate două câte două (r^2) fiind nesemnificativ.

Bateria de teste utilizată în selecţia perforatoarelor de cartele are un coeficient de corelaţie multiplă R = 0,6448, dovedind că este un bun predictor al performanţelor. În examenul psihologic al operatoarelor-perforatoare de cartele este necesară completarea analizei psihometrice cu cea clinică. În acest sens, s-a procedat la operaţia de ponderare a testelor în cadrul bateriei şi la stabilirea unui scor global ponderat. Pentru fiecare test în parte s-a acordat o anumită pondere. Scorul global a rezultat din însumarea produselor dintre rezultatele testelor şi ponderile aferente acestora.

De asemenea, au fost elaborate etaloane în sistemul staninelor care să permită o diferenţiere sensibilă a subiecţilor pentru populaţia feminină a cărei vârstă varia între 18 şi 23 de ani. Pentru estimarea eficienţei selecţiei profesionale cu ajutorul bateriei experimentale de teste, a fost întocmit un tabel de expectanţă de unde reieşea şansa de succes profesional la care ne putem aştepta.

Acest demers poate fi apreciat drept *pragmatic complex*, deoarece se referă la o baterie de teste, iar procedurile statistice care au intrat în funcţiune aparţin statisticii multivariate.

6. Tipuri de solicitări

Noţiunile de *sarcină a testului* şi *item al testului* nu se confundă. Ambele pot constitui dificultăţi care trebuie surmontate, la ambele se poate primi un anumit punctaj dacă se rezolvă situaţia problematică la care este solicitat subiectul. Bararea corectă a unui O sau C, în cadrul testului Bourdon-Amfimov, poate fi considerată o sarcină a cărei rezolvare duce la obţinerea unui punct, dar această situaţie de solicitare nu poate fi considerată un item. Deosebirea esenţială constă în faptul că itemul are autonomie informaţională, care permite efectuarea unor prelucrări statistice proprii asupra rezultatelor sale, ceea ce poate conduce la evidenţierea unor indicatori. Şi pentru rezolvarea unui item, şi pentru rezolvarea unei sarcini se poate obţine un punct, dar, pe baza analizei punctelor obţinute de diverşi subiecţi la o sarcină, nu se poate realiza o analiză statistică autonomă, deoarece am obţine valori nerelevante. Astfel, dacă încercăm să calculăm indicele de dificultate la o sarcină, obţinem frecvent valoarea 1,00 la aproape toate sarcinile.

O solicitare constă în general din două elemente:

- problema (la testele de randament şi de inteligenţă) sau întrebarea (la chestionare);
- rezolvarea problemei (la testele de randament şi inteligenţă) sau răspunsul/luarea de poziţie (la chestionare).

6.1. Tipuri de sarcini cu răspuns legat sau fixat

Răspunsurile „legate" la sarcini implică faptul că examinatului îi sunt prestabilite mai multe posibilităţi pentru răspuns, de care acesta este „legat", adică nu îşi poate permite să ofere un alt răspuns decât cel prevăzut.

• Cel mai simplu şi mai frecvent utilizat este tipul de răspuns „corect-fals", în cadrul căruia examinatul are de ales între un răspuns pozitiv şi unul negativ: între „da" şi „nu", între „exact" şi „inexact", între „corect" şi „incorect" etc. În multe cazuri, sarcina ia forma unei întrebări:

Mergeţi cu plăcere la întrunirile publice ? da ❑ nu ❑

Răspunsul corespunzător subiectului trebuie făcut cunoscut printr-o barare a variantei alese sau printr-o încercuire a acesteia (în cazul examinării unei persoane lipsite de posibilitatea de a o face, marcarea trebuie operată de către examinator). În alte cazuri, sarcina poate să ia forma unei constatări:

180 – 35 = 145 fals ❑ corect ❑
sau
Prefer locurile aglomerate, zgomotoase decât cele izolate, liniştite ?
sigur ❑ nesigur ❑

Un avantaj al unor astfel de sarcini îl constituie faptul că indicaţiile testului şi evaluarea sunt extrem de simple, iar timpul de rezolvare este scurt. Un prim dezavantaj constă în aceea că nu este întotdeauna uşor să se formuleze o întrebare sau problemă la care să se răspundă simplu prin „da" sau „nu". Se pleacă de la o ipoteză nenegociabilă, ca în cazul sarcinilor cu alegere multiplă, unde, pe lângă cel mai bun răspuns, se oferă şi alternative mai puţin bune. Există un consens în privinţa unui fapt, şi anume cu cât este mai complex conţinutul sarcinii, cu atât mai puţin se potrivesc tipurile de răspuns la aceste sarcini prin „corect/fals". Pentru sarcinile cu un conţinut complex se potrivesc modalităţile de răspuns liber sau prin completare. Alegerea de sarcini cu dispozitiv de răspuns prin „corect/fals" este considerată oportună când : 1) un test trebuie să fie rapid construit, respectiv când trebuie improvizat (desigur, ne putem întreba dacă o astfel de construcţie este un test !) ; 2) rezolvarea testului trebuie să ceară puţin timp ; 3) testul trebuie să fie evaluat de către personalul auxiliar ; 4) conţinutul sarcinii este uşor inteligibil ; 5) informaţiile despre o caracteristică a personalităţii trebuie să fie furnizate pe o bază cât mai extinsă pentru cât mai multe întrebări sau sarcini ; 6) prelucrările mintale şi randamentul memoriei nu joacă nici un rol ; 7) rezultatele întâmplătoare sunt irelevante sau sunt luate în considerare în evaluare (precauţii la evaluare). Un mare dezavantaj al tipului de sarcini al căror dispozitiv de răspuns este de tipul „corect/fals" constă în numărul mare de rezolvări întâmplătoare care pot apărea. Dat fiind acest fapt, aşteptările în privinţa fidelităţii sunt reduse, situaţie care poate să fie compensată suficient printr-o mărire a numărului de sarcini sau printr-o corecţie corespunzătoare a răspunsurilor la întâmplare (corecţie pentru care există anumite instrumente statistice). La chestionarele de personalitate pot fi utilizate cu succes tipuri de sarcină „corect/fals", cu toate că opţiunea pentru una dintre cele două nu este totdeauna uşor de făcut. Ocazional, este permisă o a treia posibilitate de răspuns (nesigur, imprecis) şi un răspuns ambivalent, cotat cu o jumătate de punct în calcularea rezultatului brut. Cele mai răspândite tipuri de sarcini sunt cele cu dispozitiv de răspuns cu alegere multiplă. Examinatul are posibilitatea de a lua la cunoştinţă mai multe posibilităţi de răspuns, de unde le poate alege pe cele pe care le consideră corecte (de exemplu, dacă este vorba despre un test de cunoştinţe) sau corespunzătoare cu propria imagine (de exemplu, dacă este vorba despre un chestionar de personalitate). Iată un exemplu :

Care dintre grupurile de litere se deosebeşte de celelalte trei :
ABCD-QRST-MWPO-WXYZ ?

Un alt exemplu este următorul :

Un comentariu este :
a) *un verdict de judecată ;*
b) *o ordonanţă ;*
c) *o lămurire ;*
d) *o concluzie ;*
e) *o stabilire.*

Sarcina, în acest caz, poate lua forma unei întrebări ca şi cea a unei stabiliri. Sarcinile cu alegere multiplă conţin de multe ori un singur răspuns corect care, în acest caz, este desemnat drept *cuvânt optim*. Dacă sarcina conţine mai multe cuvinte optime alternative, vom vorbi despre *sarcini cu mai multe răspunsuri*. Sarcinile cu alegere multiplă (*multiple choice*) sunt frecvent întâlnite la chestionarele de personalitate pentru a reliefa gradul caracteristicii personalităţii care este măsurată cu acel chestionar. În acest caz, poate fi vorba despre un răspuns în trepte ; de exemplu :

Mi se întâmplă să am o frică nejustificată !
1. *niciodată*
2. *foarte rar*
3. *rar*
4. *frecvent*

Aceste tipuri de răspunsuri mai poartă numele de *scale*, dintre care cele mai cunoscute sunt cele de tip Lickert. La evaluare, pentru fiecare răspuns este calculat un punctaj corespunzător. De exemplu : 1, 2, 3, 4 puncte sau –2, –1, 0, 1, 2. *Cuvântul optim* îşi schimbă permanent poziţia în funcţie de numărul oferit spre alegere. Există şi posibilităţi suplimentare de răspuns. Tipul de sarcină al cărei răspuns se poate alege dintre mai multe variante este predominant în construcţia testului modern. Acest tip reuneşte avantajele tipurilor de sarcină „corect/fals" şi ale celor de completare. Realizarea şi evaluarea unor astfel de sarcini sunt la fel de economice ca la primul tip de sarcină. În acelaşi timp, asemenea sarcini prezintă independenţă relativă faţă de rezolvarea întâmplătoare (aşa cum este cazul, de exemplu, la sarcinile de tip completare). Găsirea celor mai bune răspunsuri la alegere (care să constituie dispozitivul de răspuns la sarcină) ridică, în general, o serie de dificultăţi, deoarece trebuie să fie date pe cât posibil (doar) răspunsul cel mai bun şi şi alte alternative de răspuns care să apară drept plauzibile. În acest fel, în condiţiile în care examinatul nu cunoaşte rezolvările corecte, asupra acestuia se exercită o forţă de atracţie. Răspunsurile alternative trebuie să fie pe cât posibil egale ca valoare şi să aibă aceeaşi verosimilitate. Aceste condiţii trebuie permanent verificate, deoarece la analizele ulterioare se pot constata eventuale inegalităţi ale gradului de verosimilitate (de exemplu WISC). Evoluţia cunoştinţelor umane poate să reducă la una singură alternativele de răspuns, situaţie în care nu mai este realizată independenţa de rezolvările întâmplătoare. Construcţia de sarcini multiple presupune, din acest motiv, mult simţ psihologic, experienţă, conştiinciozitate, calităţi fără de care nu se pot realiza teste standardizate de înaltă competenţă. Un dezavantaj nesemnificativ al sarcinilor cu alegere multiplă constă în aceea că, prin oferirea în dar a rezolvării corecte, nu sunt evaluate suficient activitatea spontană a gândirii şi forţa acesteia. În testele de cunoştinţe, achiziţia latentă este reprodusă prin sarcinile cu alegere multiplă, iar achiziţia manifestă prin sarcinile de completare. Testele etalonate cu sarcini cu alegere multiplă au posibilităţi de aplicare dintre cele mai avantajoase, dar în două cazuri sunt mai puţin recomandate, şi anume atunci când găsirea răspunsului trebuie să se producă spontan şi atunci când pregătirea unor alternative de răspuns dă naştere la greutăţi.

• O formă de sarcină utilizată la examenele de cunoştinţe este *sarcina de atribuire*. În acest caz, sunt necesare ambele elemente ale unei sarcini: problema şi rezolvarea care se coordonează în funcţie de cealaltă. De exemplu:

Cine a scris:

1) Dona Alba	*a) Marin Preda*	*abcde*
2) Delirul	*b) Victor Ion Popa*	*abcde*
3) Păsările	*c) Geoge Călinescu*	*abcde*
4) Velerim şi Veler Doamne	*d) Gib Mihăescu*	*abcde*
5) Bietul Ioanide	*e) Alexandru Ivasiuc*	*abcde*

Referitor la ordonarea dificultăţilor, au valoare aceleaşi principii ca şi pentru întregul test, adică doar puţini candidaţi trebuie să treacă grupa individuală de sarcini fără să puncteze, la fel de puţini trebuind să rezolve toate atribuirile. Fiecare atribuire este valorizată, de cele mai multe ori, cu un punct (în exemplul nostru, maximum 5 puncte). De regulă, nu se fixează o limită temporală, deoarece majoritatea sarcinilor de atribuire sunt utilizate pentru testele de nivel. Sarcinile de atribuire sau de ordonare sunt relativ recomandabile pentru stabilirea cunoştinţelor achiziţionate. În general, au aceleaşi caracteristici bune ca şi sarcinile cu dispozitiv de alegere multiplă, mai ales când pot fi ghidate de către acestea, din punctul de vedere al teoriei testului. Sarcinile de atribuire sau ordonare se pot construi în aşa fel încât să rezulte o deplină ordonare sau atribuire a elementelor. Ultima sarcină este rezolvată de la sine, prin excluderea celorlalte sarcini rezolvate corect. Se pot introduce însă în plus una sau două sarcini neordonabile, astfel încât în fiecare grupă cu *n* sarcini să fie posibile *n–1* sau *n–2* ordonări. Avantajele speciale ale sarcinilor de atribuire constau în aceea că: 1) testul ne informează rapid despre un domeniu întins al cunoştinţelor practice; 2) un mare număr de sarcini pot să fie realizate cu un consum relativ redus de materiale.

• Pe lângă sarcinile de atribuire sau ordonare, mai există şi sarcini *de reordonare sau reatribuire*. Astfel de sarcini au însă o utilizare limitată, ca şi cele de atribuire. În cadrul acestora, i se cere subiectului să reaşeze litere, cuvinte sau figuri în mod original. Sarcinile de reordonare se utilizează la aşa-numitele *„form board tests"*, bazându-se pe principiul alegerii multiple. De exemplu, la proba WISC, subtestul „Imagini cu sens", subiectului i se cere să ordoneze imagini disparate, pentru a ilustra o întâmplare cu sens. Sarcinile de reordonare sunt mai des întâlnite la testele de prelucrare a materialelor care, în mare măsură, nu sunt standardizate şi se compun doar din puţine sarcini, uneori chiar dintr-o singură sarcină. În măsura în care astfel de sarcini se reduc la forma creion-hârtie, se iau în considerare anumite repoziţionări ale cuvintelor şi desenelor. Dacă apar două sau mai multe posibilităţi principale de rezolvare, dintre care una este calitativ mai bună, valoarea de punctaj se obţine prin ierarhizarea răspunsurilor, realizată pe baza unei analize atente. Sarcinile de reordonare ocupă mult spaţiu, când folosim forma creion-hârtie. În altă formă, este mult material de prelucrat.

6.2. Sarcini cu răspuns liber

Răspunsurile libere la sarcină sunt caracterizate prin faptul că examinatul trebuie să răspundă la sarcinile existente potrivit propriei concepţii, verbal sau nonverbal (de exemplu, prin prezentare grafică).

• Din tipurile de răspuns liber la sarcini derivă *sarcinile de completare*, precum şi cele de completare cu cuvinte-cheie (ca importanţă şi pondere, fiind apreciate după *răspunsurile cu alegere multiplă* şi răspunsurile de tip *corect/fals*). Sarcinile de completare sunt singurele forme de sarcini cu răspunsuri libere care sunt admise în testele standardizate. La aceste sarcini, subiectul este pus în situaţia de a încheia o sarcină printr-un cuvânt (cuvânt-cheie) sau printr-o scurtă prezentare. Iată un exemplu: *În ce stat se găseşte oraşul Valparaíso ?* (răspunsul corect este Chile). Sarcinile de completare sunt cele mai vechi în seria evoluţiei formelor de sarcini şi au avantajul că sunt independente de influenţa hazardului; ia naştere însă pericolul unei influenţe sugestive, când felul întrebării sau punerii problemei este legat de un anumit răspuns al examinatului. Un avantaj (şi, totodată, un dezavantaj) îl constituie împrejurarea că răspunsul este formulat după o apreciere liberă prin care se îmbogăţeşte pe cât posibil evaluarea conţinutului, însă este îngreunată evaluarea formal-statistică. Un dezavantaj mai pronunţat trebuie să fie văzut în aceea că sarcinile de completare revendică, în general, un timp considerabil mai mare decât celelalte tipuri de sarcină, fapt care face ca sarcinile de completare să se găsească rar în testele de rapiditate. Examinatul este determinat la o reflecţie temeinică, deoarece, în acest caz, nu este suficientă o sensibilitate ascuţită (ca la sarcinile de alegere multiplă) în scopul de a găsi o rezolvare corectă a sarcinii. Ca urmare a formulării libere, gradul de precizie nu poate fi decis în mod univoc (deci obiectivitatea este deficitară), motiv pentru care tipul de sarcină de completare este înlăturat dintre tipurile obiective (începând de la Binet). Tipurile de sarcină de completare sunt avantajoase în următoarele condiţii: 1) când găsirea rezolvării trebuie să se producă spontan; 2) când interesează drumul parcurs de subiect pentru rezolvarea sarcinii; 3) când trebuie să fie împiedicată ghicirea răspunsului corect sau alegerea celui mai bun răspuns; 4) când testul cuprinde puţine sarcini, complexe în privinţa conţinutului lor (de exemplu, probleme de gândire); 5) când se intenţionează, pe lângă o evaluare cantitativă, şi una calitativă (de exemplu, dacă examinatul dă, pe lângă răspunsuri false, şi răspunsuri absurde sau, pe lângă răspunsuri corecte, şi răspunsuri originale).

• Cea mai simplă formă a răspunsului la sarcini este cea a compunerilor scurte, verbale sau scrise. Sarcinile de compunere scurtă, în condiţiile în care sunt utilizate, în general, în testele obiective, trebuie să limiteze producţia subiectului la un maximum de cuvinte (de exemplu, un număr de 150 de cuvinte) în scopul de a se asigura o evaluare obiectivă a răspunsului lui. O compunere extinsă constituie o dificultate de evaluare obiectivă (de exemplu, *Cum îţi reprezinţi onoarea ?*). În afara formelor tipice, există un număr de diverse combinaţii de tipuri de sarcini individuale cu tipuri de sarcini aplicate ocazional. Astfel, sarcinile „corect/fals" se combină cu sarcinile de completare. După ce subiectul dă răspunsul „corect/fals", se permite un răspuns de completare (care, de regulă, este prezentat între paranteze). În acest mod, se pot identifica răspunsurile simple şi mai adecvate. De asemenea, se practică unificarea sarcinilor cu dispozitiv de

alegere multiplă cu cele de completare. Se cere examinaţilor ca, pe lângă alegerile dintre posibilităţile de răspuns prevăzute de sarcină, să găsească şi un răspuns mai dezvoltat, pe cât posibil mai bun, şi să-l scrie într-un spaţiu punctat. Din experienţa cu acest tip de sarcini, s-a concluzionat că răspunsurile superficiale ale subiecţilor devin mai rare. O astfel de sarcină face posibilă stimularea gândirii productive.

• Câteva teste conţin sarcini pentru care este dată o indicaţie separată, în timp ce la majoritatea are valabilitate un alt gen de recomandări, de exemplu, Army-Alfa Test, despre care am mai vorbit.

• O variantă a tipului de sarcină cu alegere multiplă o constituie sarcinile de comparaţii între aserţiuni, sarcini care au fost adaptate pentru examinarea posibilităţilor logice de decizie. De exemplu[1] :

Care dintre propoziţiile enumerate cu 1, 2, 3, 4 se află în opoziţie sau în contradicţie cu propoziţia 1 ?

1. Imperativul categoric al lui Kant este o normă morală general valabilă.

Variantele de alegere sunt :

1) *Nu toţi oamenii se conduc după imperativul kantian (nici o contradicţie) ;*
2) *Imperativul kantian este neconcludent pentru credincioşii creştini (nici o contradicţie) ;*
3) *Normele morale pot să nu existe, deoarece fiecare om este răspunzător doar în faţa conştiinţei sale (nici o contradicţie) ;*
4) *Imperativul kantian este necunoscut celor mai mulţi dintre oameni (nici o contradicţie).*

Dacă la toate cele patru sarcini din exemplul de mai sus s-a răspuns corect, se vor acorda 4 puncte. În acest caz, trebuie să spunem că la astfel de tipuri de sarcină nu se pune problema acordului sau nonacordului, ci de a se decide acordul sau nonacordul logic al constatărilor de comparat. Tipul de sarcină „comparaţii de aserţiuni" nu trebuie confundat cu tipul simplu de sarcină cu multiplă alegere.

6.3. Gradul de complexitate a sarcinii

După gradul de complexitate a conţinutului, sarcinile testului se pot împărţi în următoarele grupe :

- sarcini care pretind doar luarea de poziţie a examinatului (întrebări despre interese, dispoziţii, aprecieri). În acest caz, răspunsul la sarcină presupune un act de motivaţie ;
- sarcini care au drept premisă evaluarea unor cunoştinţe concrete (în astfel de sarcini intervine randamentul memoriei) ;

1. Adaptare după Gustav Lienert, *Testaufbau und Testanalyse*, Verlag Julius Beltz, Weinheim/ Berlin, 1967, p. 30.

- sarcinile presupunând aplicarea cunoştinţelor şi achiziţiilor (care corespund aşa-numitei *inteligenţe practice*) ;
- sarcini care presupun înţelegerea unor raporturi complexe. Rezolvarea corectă a lor este atribuită aşa-numitei *inteligenţe teoretice.*

Se poate observa că o astfel de diviziune se bazează pe un principiu de ordonare abstract şi nu are nici o importanţă practică.

În concluzie, putem afirma că în examinarea psihologică nu se recomandă folosirea exclusivă a unui tip de sarcină. Practic, la un test simplu se foloseşte un singur fel de sarcină, iar într-un sistem de teste trebuie să fie schimbat tipul de sarcină de la un subtest la un alt subtest.

Ordonarea tipurilor de sarcină ar trebui să fie astfel întreprinsă încât :

- subtestele care presupun o mare concentrare (respectiv tipurile de sarcină corespunzătoare) să fie programate cu preponderenţă la începutul testului ;
- sarcinile uşoare de tipul „corect/fals" şi cele dificile de tip completare să alterneze.

Cu privire la ordonarea conţinutului sarcinii (gândire, memorie, inteligenţă) şi a tipurilor de sarcină corespunzătoare, nu s-a stabilit nici o regulă. În mod principial, conţinutul unei sarcini poate să îmbrace orice formă, ceea ce este valabil, înainte de toate, pentru sarcinile de inteligenţă. Pentru chestionarele de personalitate, sunt utilizate predominant tipurile de sarcină „corect/fals" şi cu dispozitiv de răspuns cu alegere multiplă. La chestionarele de atitudini şi opinii, sunt utilizate, în mod special, tipuri de sarcină cu alegere multiplă şi tipuri de sarcini cu răspuns în trepte.

7. Problema eşantionării

Construcţia unui test necesită examinarea preliminară a unui număr apreciabil de subiecţi. La realizarea MMPI au fost examinaţi, numai în faza de început, 700 de subiecţi dintr-un eşantion de locuitori ai statului Minnesota, 350 de candidaţi la colegiu şi 800 de pacienţi ai spitalului din Minneapolis. Grupa de standardizare pentru spaţiul lingvistic german a cuprins 365 de persoane. Pentru a exclude posibile deformări ale sensului întrebărilor, prin traducerea în germană a MMPI, acest chestionar a fost aplicat pe un număr de 700 de bilingvi (Hathaway şi McKinley, 1963).

Eşantionările în psihologie se constituie mult mai greu decât în alte domenii, cum ar fi sociologia. De exemplu, psihologii nu pot merge decât în rare cazuri la domiciliul subiecţilor pentru a face investigaţii ştiinţifice. În marea majoritate a cazurilor, examenul psihologic se desfăşoară în spaţii special amenajate, cu respectarea anumitor etape. Chiar dacă anumiţi subiecţi reuşesc să răspundă unei invitaţii pentru a fi examinaţi într-un laborator psihologic, nu putem fi siguri că au o motivaţie corespunzătoare pentru efectuarea examinării. În ultima vreme, începe să se practice plata subiecţilor, încât problemelor de eşantionare li se adaugă unele noi, legate de dimensiunea financiară a unei examinări.

Practic, subiecţii care pot fi examinaţi în condiţii de asigurare a ambianţei specifice examinării psihologice sunt cei selectaţi prin intermediul unei instituţii (pacienţii din spitale, elevii din licee, persoanele care vin la centrele de evaluare sau la laboratoarele unor mari întreprinderi ori regii naţionale). În momentul în care un subiect ştie că de

rezultatul examenului depinde viitorul său profesional, există o mare probabilitate să răspundă motivat la solicitările testului. Considerăm că subiecţii folosiţi la verificarea probelor psihologice sunt, astfel, selectaţi în mod implicit. Constituirea unor eşantioane complet aleatorii nu este imposibilă, dar duce la complicaţii mari de ordin metodologic.

În privinţa eşantionării, avem două situaţii distincte: eşantionare de persoane care efectuează examinarea şi eşantionare dintr-un anumit univers specific de itemi. În cazul primei situaţii de eşantionare dintr-o anumită populaţie, este vorba despre o eşantionare probabilistă (prin tragere la sorţi elementară, prin tragere la sorţi secvenţială, prin *grappes* şi stratificare) sau despre una empirică (prin *quote*, prin unităţi-tip). În privinţa eşantionării itemilor, există eşantionare simplă, eşantionare stratificată, eşantionare prin *grappes* şi eşantionare ierarhică.

Chiar în situaţia în care subiectul este condiţionat instituţional, apar o serie de probleme, una dintre ele fiind compoziţia unui grup-criteriu. Dacă în unele situaţii problema este simplă (cercetarea în psihologia şcolară, de exemplu), în psihologia patologică nu există un adevărat acord între specialişti în materie de diagnostic. Existenţa unor nomenclatoare, de genul DSM IV, nu rezolvă problema constituirii unor grupe de subiecţi-criteriu. Paul Kline (1995, p. 120) spune în această privinţă: „Chiar dacă este posibil să fie selectaţi plecând de la un criteriu pentru a constitui un grup, subiecţii repartizaţi acestui grup-criteriu pot să se diferenţieze de subiecţii grupului-martor pe mai mult de o variabilă. Nişte subiecţi schizofreni (repartizaţi grupului-criteriu) pot să se diferenţieze de subiecţii grupului de control în domeniul personalităţii, al funcţiei sociale şi al aptitudinilor intelectuale, pentru a nu reţine decât trei variabile importante".

8. Normarea rezultatelor la testele psihologice

Cuvântul „normă" are o arie extinsă şi diversificată de utilizare, cuprinzând domeniile disciplinelor ştiinţifice şi ale activităţilor practice. Se vorbeşte despre norme juridice, norme etice, norme deontologice, norme de circulaţie, norme de consum, norme tehnologice, norme ecologice, norme de comportament, norme de reprezentare etc.

Sensul general al termenului de *normă*, în disciplinele psihologice aplicative, este acela de ansamblu de date caracteristice unei populaţii, pentru o sarcină de efectuat şi un material determinat, pe care un cercetător îl utilizează în momentul construcţiei unui plan experimental. Norma deţine un loc deosebit de important în psiholingvistică şi în psihologia cognitivă. Normele lingvistice sunt constituite din frecvenţele medii de apariţie a unui cuvânt în limbă, acestea stabilindu-se după analiza unui număr de texte variabil din punct de vedere calitativ. S-a putut stabili că datele lingvistice sunt bine corelate cu estimarea subiectivă directă a *familiarităţii cuvintelor.*

La începutul cercetărilor psiholingvistice, s-au determinat *normele asociative*, care indicau pentru un ansamblu de cuvinte răspunsul dat cel mai frecvent de fiecare subiect, cu titlu de asociaţie liberă. Această indicaţie este însoţită de precizarea procentajului subiecţilor care au dat respectivele răspunsuri.

Normele sunt, totodată, convenţii care reglează viaţa socială şi trebuie să rezolve conflictele înainte ca ele să izbucnească. Normele de comportament, în calitate de „prescripţii nescrise" constituite de un grup sau o societate, preocupă mai ales psihologia socială. Interesează, în special, caracteristicile, geneza şi schimbarea normelor, sistemul normativ şi devianţa comportamentală etc.

Cel mai frecvent, cuvântul „normă" se foloseşte în domeniul psihometriei, parte integrantă a psihologiei diferenţiale, disciplină care se ocupă cu teoria şi tehnologia instrumentelor de măsură în psihologie. Deseori, acest termen se foloseşte în strânsă legătură cu cel de *etalon*, iar cel de *normare* cu cel de *etalonare*. A etalona un test şi a norma un test sunt considerate activităţi identice. În definiţia lui Maurice Reuchlin, personalitate ştiinţifică de referinţă în domeniul utilizării metodelor cantitative în cercetarea psihologică, etalonarea este „stabilirea, într-o probă psihologică, a unei scale permiţând de a situa rezultatul obţinut de un subiect prin raportare la rezultatele care au fost observate anterior într-o populaţie de referinţă, suficient de numeroasă şi omogenă, de subiecţi comparabili aceluia care a fost examinat"[1]. Etalonul, în general, este o mărime standardizată servind drept referinţă şi măsură între diverse obiecte. A etalona înseamnă a stabili gradaţiile pentru un instrument de măsură sau pentru o scală. A norma înseamnă a fixa pentru o mărime numărul de norme sau de etaloane care sunt cuprinse. Atât etalonul, cât şi norma presupun o convenienţă între utilizatorii de măsuri. Considerăm că normarea are un grad de convenienţă crescut în comparaţie cu etalonarea, operaţiune care implică un interes crescut în a asigura identitatea semnificaţiei unităţii de măsură. Atunci când efectuăm o operaţiune de măsurare, depindem foarte mult în apreciere de puterea informaţională pe care am dobândit-o în urma efectuării acesteia. Nu ne este indiferent dacă ne plasăm în domeniul scalar nominal, ordinal, de log-interval, de interval, de raport sau absolut. Ignorarea acestui fapt duce la grave erori de aprecieri şi la operaţii statistice lipsite de sens.

Problema normării rezultatelor cantitative în cercetarea psihologică a apărut o dată cu dezvoltarea psihologiei experimentale şi, mai ales, cu apariţia testului psihologic, dar rămâne încă de actualitate. Această problemă apare frecvent în interpretarea cantitativă a datelor provenind din aplicarea instrumentelor psihologice de investigaţie şi în docimologie, în general. Normarea poate fi aplicată pentru orice metodă psihologică de investigaţie ale cărei rezultate pot fi exprimate mai ales cantitativ, dar în special în domeniul construcţiei şi aplicării testelor psihologice, care sunt instrumente standardizate. În general vorbind, standardizarea procedurilor este indispensabilă comunicării ştiinţifice a informaţiilor, care se prezintă sub formă cantitativă, şi oricărui demers metodologic de comparare. Ea este inerentă oricărei observaţii ştiinţifice.

8.1. Problema modelelor de măsură

A aplica un model la serii de date rezultând din cercetare presupune aderarea la o concepţie asupra fenomenului studiat. Orice test construit conform rigorilor ştiinţifice are la bază o teorie asupra caracteristicii propuse spre diagnosticare, teorie care poate fi implicită sau explicită. În psihologie, un model este o particularizare a unei teorii de funcţionare cognitivă. Când un model este pus în aplicare, sunt testate numeroase ipoteze, şi anume acelea conţinute în formalizarea sa. Toate modelele de măsură se aplică la datele de intrare şi conduc la nişte măsuri rezultante, care pot fi puse în legătură cu alte măsuri, pentru a studia relaţiile interconcepte. Aceste măsuri pot fi de diferite niveluri, după cum urmează:

- nivelul nominal, în cazul analizei în clase latente;

1. Vezi *Grand dictionnaire de la psychologie*, Larousse, Paris, 1992, p. 284.

- nivelul ordinal, în cazul analizei scalogramei individuale a lui Leon Guttman;
- nivelul de interval, pentru modelul judecăților comparative sau pentru modelele de răspunsuri la itemi.

Măsurile suportă toate transformările permise de nivelul lor. Transformările sunt realizate din nevoia unei mai simple lizibilități a rezultatelor sau din cea de comunicabilitate. Dacă sunt perfect legitime, aceste transformări pot aduce o semnificație suplimentară, pot facilita lectura rezultatelor și compararea acestora (de exemplu, compararea rezultatelor copiilor de vârste diferite).

Există trei mari abordări care permit transformarea măsurilor în informații direct interpretabile (acestea nu sunt concurente). Ele se disting prin tipul de informații aduse și cadrul de referință ales.

1) Dacă se alege *cadrul de referință predictiv*, se stabilește o corespondență între măsurile originare și alte măsuri, în acest fel putându-se face predicții asupra celor din urmă.
2) În cazul în care se alege drept cadru de referință *domeniul*, măsurile se vor interpreta în termeni de măsură-criteriu specifică domeniului. Acum, subiecții se împart în cei care depășesc, prin randamentul obținut, această măsură-criteriu și cei care nu o depășesc. Situarea de o parte sau de cealaltă a respectivei „bariere" face posibile afirmații asupra valorii, competenței, reușitei în domeniu. Nu vom aborda detaliat problematica privind constituirea unei măsuri-criteriu.
3) Cadrul de referință *normativ* permite să se situeze măsura unei caracteristici prezente la un subiect prin raportare la măsurile aceleiași caracteristici prezente la subiecți comparabili. Se vorbește, în acest caz, despre etalonare. În cadrul de referință dat, se pot distinge mai multe tipuri de transformări.

8.2. Metode de etalonare a rezultatelor

Interpretarea scorurilor obținute la un anumit test psihologic se face în etape. Prima etapă o constituie transformarea scorurilor brute într-o scală care permite stabilirea unei baze de comparație.

Orice subiect căruia i-a fost aplicat un instrument de investigație psihologică obține un rezultat observabil, exprimat conform prescripțiilor manualului de răspunsuri, în urma aplicării grilei de corecție. Testul computerizat exclude existența unei grile de corecție ca material al testului, procedând automatizat, având o precizie și o rapiditate incomparabile cu modalitățile tradiționale de corectare. Redarea automată a rezultatelor brute și a diferitelor feluri de transformări posibile nu exclude necesitatea învățării modului de constituire a etaloanelor și a cunoștințelor privind reconsiderarea și reevaluarea lor. Operațiile tehnice care stau la baza unei astfel de practici nu trebuie totuși omise de către cercetători și practicieni.

Testele de randament intelectual și cele pentru măsurarea diferitelor abilități psihomotorii au, în general, răspunsurile împărțite în corecte și incorecte. Realizarea punctării unui item individual este în concordanță cu valoarea răspunsului: 0 puncte pentru răspuns incorect, 1 punct pentru răspunsul corect. La inventarele de personalitate și de interes, răspunsurile nu se apreciază în acești termeni, ci în funcție de evidențierea sau nu a caracteristicii investigate sau a atitudinii exprimate față de conținutul aserțiunii unui item.

În majoritatea cazurilor, rezultatele aplicării unor instrumente de investigație psihometrică dobândesc consistență cantitativă prin operațiile de adiție realizate sub formă matriceală. Cel mai simplu și mai frecvent tip de matrice este *matricea lui W. Stern*, propusă de psihologul american în 1912 și numită tradițional *matricea itemi – subiecți*.

O astfel de matrice de date poate fi reprezentată pe două dimensiuni:

- liniile, care reprezintă răspunsurile unui același subiect la itemi diferiți;
- coloanele, care cuprind rezultatele diferiților subiecți la un același item.

Deoarece elementele liniilor sunt constituite din rezultatele subiecților la fiecare item, din analiza acestora putem trage concluzii asupra unor profiluri de răspuns ale subiecților (cărora li se spune și *pattern*-uri de răspuns). Din analiza coloanelor, putem trage concluzii asupra calității itemilor. Indici de similaritate și de distanță pot fi calculați atât pe liniile, cât și pe coloanele aceleiași matrice. În 1966, R.B. Cattell a propus o extensie a matricei lui W. Stern pentru a veni în întâmpinarea unei practici curente a cercetării, constând în a aprecia stabilitatea temporală sau intersituațională. Dacă se repetă același chestionar în mai multe ocazii, se obține o *matrice subiecți – itemi – ocazii*. Ocaziile temporale vor ocupa o a treia intrare a unei structuri tridimensionale, numită în mod obișnuit *cutia lui Cattell*.

Trebuie să reamintim că o serie de teste psihologice de randament psihomotoriu nu sunt compuse din itemi, ci punctează diferite criterii de apreciere a îndeplinirii sarcinii (în acest caz, punctarea este constituită de numărul de erori, durata medie a acestora, timpul de efectuare a probei etc.).

Problema privind măsurarea subiecților după o trăsătură, o caracteristică sau o abilitate se rezumă la însumarea răspunsurilor la un anumit număr de itemi în așa fel încât să se obțină un rezultat care să fie rezultanta măsurii pentru fiecare persoană. O asemenea însumare dă naștere la un *scor*. Operațiunea de creare a unui scor este numită de specialiștii psihometricieni *scorare*[1]. Scorul realizat prin adăugarea valorilor obținute la itemi poartă numele de *scor compozit*. Itemii se pot prezenta și mai diversificat decât notarea dihotomică, situație întâlnită mai ales la inventarele de atitudini. Aproape întotdeauna se pot reprezenta modele de măsurare a indivizilor, prin trasarea unei curbe care exprimă grafic legătura dintre intensitatea caracteristicii și probabilitatea de a obține un răspuns corect. Chiar dacă nu suntem niciodată siguri de faptul că măsurăm bine o anumită caracteristică, se poate aprecia că scorurile obținute dintr-o combinație de itemi constituie o bună aproximare.

Scorurile brute sunt acele scoruri obținute direct, imediat după corectarea probei, fără nici o altă intervenție prealabilă. Aceste variabile brute au amplitudini foarte diferite, ceea ce face să nu ne putem orienta apreciativ în privința locului relativ al unui subiect într-un ansamblu de subiecți care au efectuat o anumită probă. Scorul 75, la un test psihologic permițând obținerea de note brute de la 0 la 200, nu poate fi comparat cu cel de 178, obținut la un test unde se pot obține note între 0 și 380. Pentru a putea efectua comparații între scoruri, este nevoie de o serie de prelucrări matematice ale valorilor brute.

Un scor brut nu are semnificație psihodiagnostică decât dacă este pus în legătură cu randamentul unui grup de respondenți comparabili. Grupul de referință se mai cheamă și *grup normativ*. Primul element ce trebuie luat în seamă, în cazul în care dorim să

1. Noțiunea este folosită frecvent de Constantin Mitrofan, în lucrarea *Testarea psihologică a copilului mic*, Editura All, București, 1997.

aflăm poziţia relativă a unui subiect în interiorul unui grup, îl reprezintă tabelul cu norme referitoare la ansamblul randamentelor individuale din grupul de referinţă la un anumit instrument psihometric de investigaţie. Utilizarea normelor este ghidată de unele principii de interpretare a rezultatelor. Respondentul al cărui scor este comparat cu norma sau etalonul trebuie să posede anumite caracteristici asemănătoare acelora ale grupului de apartenenţă. Măsurarea cu ajutorul unui test psihologic, construit în spiritul teoretic clasic, capătă sens doar prin raportare.

Pentru a putea raporta valorile brute obţinute de un subiect la valorile de referinţă, trebuie efectuate examinări pe grupuri de subiecţi posedând aceleaşi caracteristici. Este necesar să se facă distincţia între noţiunea de *populaţie* şi cea de *eşantion*. *Populaţia* este constituită din ansamblul indivizilor (în domeniul psihologic, ansamblul subiecţilor) care satisface exigenţele unei anumite definiţii. Grupul de referinţă poartă numele de *eşantion* doar atunci când selecţia subiecţilor s-a făcut prin respectarea rigorilor unor anumite reguli, fapt ce-i conferă acestuia reprezentativitate (calitatea care face posibilă tragerea aceloraşi concluzii din analiza grupului ca şi din analiza populaţiei, admiţând anumite riscuri tolerabile de eroare).

Procedura de etalonare a unui test cuprinde efectuarea acestuia de către cel puţin un eşantion din populaţia de referinţă. Începerea procedurii trebuie să se facă printr-o definire foarte clară a acestei populaţii. Din punct de vedere teoretic, ansamblul poate fi finit sau infinit, dar datele cu care se confruntă psihologul sunt în toate cazurile în număr finit. Dacă populaţia este constituită dintr-un număr redus de elemente, ea poate fi măsurată în întregime. Dar colectarea exhaustivă de răspunsuri ale subiecţilor din populaţie este, de cele mai multe ori, imposibilă. Trebuie să ne limităm la domeniul eşantioanelor, adică la părţi, subansambluri ale populaţiei, plecând de la care caracteristicile populaţiei trebuie aproximate într-un mod satisfăcător.

În definirea populaţiei, trebuie să se ţină cont de o serie de caracteristici precizate de destinaţia psihodiagnostică a testului. Dacă testul este destinat a diagnostica tulburările de dezvoltare senzorio-motorie a copilului mic (0-2 ani), populaţia vizată va fi, desigur, aceea a copiilor în vârstă de la 0 la 2 ani. Un chestionar care îşi propune să evalueze dezvoltarea socială a handicapaţilor mintal va avea ca populaţie de referinţă această categorie demografică defavorizată. O recomandare generală este ca populaţia de referinţă să fie suficient de omogenă, adică din ea să facă în mod clar parte toţi indivizii susceptibili de a fi evaluaţi cu ajutorul unui test. Când un test este dezvoltat de către un editor comercial, există frecvent exigenţa ca normele să fie naţionale. Avantajul major de a se referi la o populaţie naţională este de a permite producţia unui sistem unic de norme, valabil pentru un foarte mare număr de subiecţi. Interesul comercial şi uşurinţa de folosire sunt evidente. Referinţa la populaţia naţională nu presupune ca normele diferitelor teste să fie *ipso facto* comparabile, deoarece această populaţie nu este întotdeauna definită cu aceeaşi rigoare de către cei care editează un instrument de investigaţie. Neconcordanţele survin atunci când editorii nu se pun de acord în legătură cu includerea anumitor grupuri atipice în populaţia de referinţă. Deficienţii mintal sunt câteodată incluşi, iar alteori excluşi din populaţie. Ca atare, normele naţionale sunt deficitare, datorită prea extinsei lor generalităţi. De aceea, este mai adecvat ca practicienii să ia decizii sprijinindu-se pe norme specifice. Un psiholog care activează profesional în medii socioeconomice defavorizate va avea mai mare nevoie, în general, de norme elaborate pentru respectivul tip de populaţie. Această raţiune, alături de cea financiară, este frecvent invocată ca justificare a practicii de a nu dezvolta decât norme locale. În acest caz, populaţia de referinţă va fi mai bine circumscrisă. Ea va corespunde, de exemplu, elevilor din şcolile unui oraş sau pacienţilor dintr-o instituţie de recuperare psihomotorie. Normele vor avea referinţă la aceste populaţii şi vor servi, în mod obişnuit, pentru obiective foarte precise: sprijinirea orientării elevilor către diferite profesiuni, constituirea de grupuri omogene pentru învăţare etc. Limitele folosirii normelor locale decurg din prea marea lor specificitate. În acest caz, pentru alte utilizări ale testului sau chestionarului, va fi adesea necesar să se dezvolte noi norme, extinse în privinţa ariei de apartenenţă a populaţiei vizate.

Calculele efectuate pe eşantioane nu constituie decât o estimare a anumitor parametri ai populaţiei, ca media şi varianţa scorurilor. În statistică, noţiunea de *parametru* are diferite sensuri. Unul dintre ele este acela de mărime statistică fixă, rezultată din prelucrarea rezultatelor la investigarea întregii populaţii. Un alt sens este acela de mărime statistică obţinută din date care îşi conformează perfect distribuţia exigenţelor unui model teoretic (de exemplu, Gauss-Laplace). Scopul procedurii de eşantionare este de a minimiza eroarea de estimare a acestor parametri, luaţi în primul sens amintit.

Din raţiuni de economie, se procedează frecvent la o eşantionare de convenienţă, mai comodă pentru practician, utilizând subiecţi apropiaţi anturajului psihologului sau persoane care se prezintă voluntar în urma unui anunţ. Această procedură trebuie respinsă, deoarece antrenează serioase deformări sistematice în estimarea parametrilor populaţiei. Lăsarea la aprecierea practicianului nu conduce, în general, la constituirea unui eşantion reprezentativ din populaţie, deoarece erorile datorate lipsei de sistematizare sunt foarte dificil de controlat. Mai mult, procedura nefiind aleatorie, nu este posibil să se evalueze importanţa erorii de estimare a parametrilor. În 1971, W.H. Angoff face precizarea că, pentru testele cognitive, folosirea eşantionării de convenienţă conduce în mod obişnuit la o supraestimare a scorurilor populaţiei. Într-adevăr, subiecţii voluntari sau care fac parte din anturajul cercetătorului constituie un subgrup sociocultural favorizat în rândul populaţiei. Exemplul cel mai frecvent dat în privinţa deformării rezultatelor din sondaje este cel al erorii de eşantionare de la alegerile prezidenţiale din Statele Unite, din anul 1948, când toate marile agenţii de sondare a opiniei publice au prevăzut victoria lui Thomas E. Dewey contra lui Harry Truman. Eroarea, în acel caz, s-a datorat folosirii eşantionării prin *quote*. Slăbiciunea respectivei metode de eşantionare constă în a ignora intervenţia subiectivităţii intervievatorului. O preferinţă de natură inconştientă riscă să intervină în selecţia respondenţilor. Pentru controlarea erorii de eşantionare, adică a erorii de estimare a parametrilor populaţiei, este necesar să se excludă orice subiectivitate de procedură şi să se constituie eşantioane în manieră pur aleatorie. Un eşantion poate fi considerat aleatoriu dacă fiecare subiect din populaţie are o probabilitate egală de a fi selectat. În anumite cazuri, estimarea parametrilor populaţiei poate fi lipsită de influenţa factorilor de eroare. Este posibil să se calculeze eroarea standard a estimării diferiţilor parametri şi să se determine un interval de încredere în jurul valorilor calculate, plecând de la scorurile eşantionului. Există diverse tehnici de eşantionare aleatorie: 1) eşantionare pur aleatorie, 2) eşantionare aleatorie stratificată; 3) eşantionare sistematică şi 4) eşantionare prin *grappes*.

Vorbim despre eşantionare aleatorie simplă atunci când, dintr-o populaţie de mărime N, extragem un eşantion de mărime n, în aşa fel încât fiecare individ din populaţie să

aibă aceeaşi probabilitate de a fi selectat. Procedura de eşantionare aleatorie simplă constă în a asigura un număr specific fiecărui individ din populaţie, apoi în a trage la sorţi dintre toate numerele un eşantion a cărui mărime să fie stabilită în prealabil. Pentru a realiza această tragere la sorţi, utilizăm fie o tabelă de numere aleatorii, fie un generator de numere aleatorii, inclus în majoritatea programelor de statistică actuale (SPSS, SAS, SYSTAT, SigmaStat). Tabelele de numere aleatorii sunt construite în scopul realizării unei distribuţii uniforme. Programele statistice permit şi generarea de numere aleatorii cu realizarea diferitelor tipuri de distribuţie (distribuţie *z*, distribuţie *t*, distribuţie $\lambda 2$ etc.). În eşantioanele pur aleatorii, va fi necesar să se aleagă procedura de generare a numerelor aleatorii cu o distribuţie uniformă, indivizii având astfel şanse egale de a fi aleşi.

După stabilirea eşantionului, trebuie să procedăm la aplicarea testului, pe care dorim să-l etalonăm, pe subiecţii respectivi. Pe baza scorurilor eşantionului, vom calcula statisticile de care suntem interesaţi, considerate ca fiind estimări ale parametrilor populaţiei. Erorile de eşantionare fiind inevitabile, are importanţă în mod egal evaluarea erorii de estimare a parametrilor. Media pe care o calculăm plecând de la scorurile eşantionului riscă să fie sensibil diferită de aceea pe care o putem calcula plecând de la scorurile populaţiei. Dacă extragem un mare număr de eşantioane din rândul populaţiei şi calculăm de fiecare dată media scorurilor, valorile diferitelor medii tind să se distribuie normal şi media mediilor lor va fi egală cu media populaţiei. Abaterea standard a acestei distribuţii a mediilor se va numi *eroarea standard a mediei* şi se va nota cu s_M. Plecând de la un eşantion selectat, această valoare poate fi estimată în manieră nedeformată prin intermediul formulei:

$$s_M = \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)},$$

unde s^2 = varianţa scorurilor eşantionului, n = mărimea eşantionului, iar N = mărimea populaţiei.

În această formulă, după cum precizează Dany Laveault şi Jacques Grégoire (1997, p. 265), cantitatea $(N-n)/N$ se numeşte *corecţie a populaţiei finite*. Această corecţie ia în considerare faptul că o estimare bazată pe un eşantion de 20 de subiecţi extras dintr-o populaţie de 60 de subiecţi conţine mai multă informaţie în legătură cu populaţia decât un eşantion de 20 de subiecţi extras dintr-o populaţie de 10.000 de subiecţi. Această corecţie poate fi ignorată când este superioară sau egală cu 0,95, adică $n \leq (1/20)N$. În acest caz, formula se scrie într-un mod mai simplu:

$$s_M = \sqrt{\frac{s^2}{n}}$$

Cunoaşterea erorii medii ne permite să construim un interval de încredere în jurul mediei eşantionului. Acest interval ne ajută să estimăm valoarea mediei populaţiei plecând de la eşantion şi să luăm cunoştinţă de importanţa erorii de estimare a mediei.

8.3. Tipuri de scale de măsură în psihometrie

În general, în psihometrie se face raportare la cinci categorii de scale[1]:

- scale de conţinut;
- scale de dezvoltare;
- scale care utilizează raporturi şi coeficienţi;
- scale care se bazează pe cuantumuri procentuale sau cuantile;
- scale care fac apel la scoruri standardizate.

8.3.1. *Scalele de conţinut*

Aceste scale permit raportarea randamentului unui respondent la un standard definit prin conţinutul testului. Ele se folosesc mai ales la acele teste specifice în activitatea de selecţie profesională, în care sarcina este foarte apropiată activităţii reale pe care o va avea de efectuat candidatul dacă va fi admis. Exemplul clasic este proba maşinii de scris. În acest caz, este stabilit clar un randament maximal, ideal, de care subiectul se poate apropia şi pe care, în cazuri excepţionale, îl poate chiar realiza. Se compară deci un randament individual cu un randament ideal. Acest tip de scală nu se aplică decât la testele de randament maximal. Randamentul subiectului se calculează cu următoarea formulă:

$$R = \frac{\text{Numărul de răspunsuri corecte}}{\text{Numărul total de itemi}} \times 100.$$

Dacă subiectul ar fi putut realiza un randament maxim de 400 de puncte şi nu a realizat decât 324, atunci se va face următorul calcul: $\frac{324}{400} \times 100 = 81\%$. Scorurile derivate în acest mod depind foarte mult de proprietăţile itemilor din test. Un scor de 81% va fi interpretat după cum itemii testului sunt uşori sau dificili. Pentru doi respondenţi, este posibil să se obţină acelaşi scor, cu răspunsuri corecte la itemi comportând un grad de dificultate variabil. De exemplu, subiecţii A şi B obţin amândoi 70%, dar A a reuşit doar itemi uşori, iar B a reuşit şi itemi dificili. De aceea, nu putem spune că performanţele celor doi subiecţi sunt de acelaşi nivel.

Scalele de conţinut pun o serie de probleme:

- validitatea trebuie să fie asigurată printr-o eşantionare adecvată bazată pe o definiţie precisă a domeniului măsurat;
- interpretarea scorurilor trebuie să se facă în funcţie de conţinutul măsurat şi de itemii rezolvaţi de respondent.

Asemenea scale prezintă avantajul că permit interpretarea scorurilor calculate în funcţie de randament. Aceste scoruri indică gradul de abilitate pe care îl posedă în prezent un respondent, mai curând decât poziţia sa relativă într-un grup.

Dezavantajele constau în dificultatea de a preciza conţinutul şi natura sarcinilor complexe şi în dificultatea de a afirma că un respondent stăpâneşte un nivel de dificultate

1. O scală variază după postulatele subiacente la fiecare nivel de măsură. Astfel, centilele sunt ranguri; scorurile standard sunt bazate de o unitate standard *z* şi constituie o scală de intervale egale. Calculul coeficienţilor nu este posibil decât plecând de la scale de proporţii.

dat pentru că a reuşit, probabil, atât la itemi uşori, cât şi la itemi dificili. Interpretarea unui scor în sine este întotdeauna aproximativă, deoarece noţiunea de dificultate se raportează la un grup de subiecţi mai mult decât la un singur subiect.

8.3.2. *Scale de dezvoltare*

Există două tipuri de scale de dezvoltare, în funcţie de vârstă şi de nivelul de şcolarizare. Principiul care stă la baza construcţiei acestora este că anumite abilităţi şi caracteristici se dezvoltă într-un mod sistematic. Scalele de dezvoltare se utilizează pentru a compara randamentul unui subiect cu randamentul mediu al altor subiecţi care au atins un anumit nivel de dezvoltare.

Rezultatele măsurate pe scale de dezvoltare sunt tributare unui număr de factori, precum educaţia, mediul socioeconomic, influenţa culturală, stabilitatea emoţională etc. Acest tip de scală pare să convină în cazul copiilor dintr-un mediu „tipic".

8.3.3. *Scala pe niveluri de vârstă*

Postulatul fundamental al acestei scale de dezvoltare este acela că gradul prezenţei unor abilităţi sau caracteristici măsurate creşte o dată cu vârsta. Dacă o caracteristică nu sporeşte pe măsura înaintării în vârstă, scala nu este adecvată pentru măsurarea acesteia. Dacă procentul de schimbare variază de la an la an, inegalitatea mărimii unităţii complică interpretarea. În general, dezvoltarea mintală se stabilizează la adolescenţă.

În raport cu vârsta, Alfred Binet a considerat că dezvoltarea mintală se estimează prin compararea randamentului unui copil cu acela al copiilor de vârste diferite. El a selecţionat sarcini intelectuale (itemi) care îi permiteau să discrimineze copii de vârste diferite. De exemplu, itemul 9 este efectuat de doar câţiva copii de 7 ani şi de majoritatea celor de 8 ani. În concluzie, el este recomandat pentru vârsta de cel puţin 8 ani. Binet a realizat colecţii de itemi pentru examinarea fiecărei vârste. Scorul reprezentativ este cel al eşantionului pentru o anumită vârstă.

Vârsta mintală este vârsta corespunzătoare performanţelor realizate de către copiii normali de o anumită vârstă, la anumite solicitări de natură intelectuală. Este determinată prin nivelul probelor reuşite de către un anumit copil. Probele create de Binet (şi, ulterior, de Simon) constituie o scală ierarhică. Vârsta mintală constituie nivelul de vârstă pentru care toate răspunsurile specifice sunt corecte.

În 1905, A. Binet şi Th. Simon prezintă un set de treizeci de probleme în ordine crescătoare. Nivelul de dificultate a fost determinat empiric, administrând probele unui număr de 50 de copii normali în vârstă de la 3 la 11 ani şi la câţiva copii şi adulţi retardaţi. Nu este dată nici o metodă de a ajunge la un scor total. În 1908, numărul probelor a crescut, unele fiind eliminate; ele sunt regrupate pentru nivelul de vârstă mintală în funcţie de rezultatele obţinute pe 300 de copii normali. După 1911 (an în care testul modificat comportă cinci itemi pe nivel de vârstă), vârsta mintală se calculează ţinând cont de reuşita parţială la vârstele superioare vârstei de bază. Astfel, un copil care a reuşit toţi itemii pentru 7 ani şi cinci itemi pentru un nivel superior va avea o vârstă mintală avansată în funcţie de reuşita suplimentară. S-a constatat că un acelaşi număr de ani de avans sau de întârziere nu are aceeaşi semnificaţie după vârsta reală. Un an de întârziere la 4 ani nu are aceeaşi semificaţie ca un an de întârziere la 11 ani.

Pentru a ţine cont de această situaţie, William Stern a introdus, în 1912, noţiunea de QI, desemnând raportul între vârsta mintală şi vârsta cronologică (sau vârsta reală calculată în luni). Acest raport este înmulţit cu 100 şi se elimină zecimalele. Un QI de 100 reprezintă o performanţă potrivită cu vârsta reală, unul inferior lui 100 indică o „întârziere" în viteza de dezvoltare, iar un QI mai mare decât 100 indică un „avans" în dezvoltare.

Constituirea de scale pe niveluri de vârstă se derulează în mai multe etape:

1. Sunt constituite eşantioane de subiecţi pentru o anume vârstă. O vârstă este definită ca un interval mai mult sau mai puţin extins decât vârsta în chestiune.
2. Este calculat un scor mediu la fiecare grupă de vârstă.
3. Eventual, scorurile diferitelor grupe de vârstă sunt estimate prin interpolare. Procedura se utilizează când anumite vârste nu au fost incluse în eşantionul de etalonare şi când se doreşte prezentarea normelor pe luni, nu numai pe ani. Procedura de interpolare răspunde la postulatul unei progresii liniare a caracteristicilor evaluate prin test. Ea constă în a calcula valoarea intermediară între valorile inventariate din eşantionul de etalonare. De exemplu, dacă scorul mediu obţinut la test pentru copiii de 7 ani este de 14 puncte şi cel obţinut de copiii de 9 ani este de 18 puncte, se poate estima că scorul mediu al copiilor de 8 ani este de 16 puncte.

8.3.4. *Scale de nivel şcolar*

Se compară randamentul şcolar al unui subiect cu acela al unui cursant mediu din diverse niveluri şcolare. De exemplu, 7,5 este nivelul mediu de 7 ani şi 6 luni, adică un subiect cu acest scor are acelaşi randament cu un şcolar mediu de 7 ani şi 6 luni. Stabilirea valorii respective indică faptul că elevul are nivelul corespunzător de cunoştinţe. Este vorba în acest caz despre o măsură continuă. Trăsătura trebuie să varieze în raport cu nivelul şcolar. Efortul făcut pentru a determina şi reevalua nivelurile şcolare este foarte mare şi cercetătorul se loveşte de nenumărate dificultăţi pe care nu le poate depăşi decât stabilind o serie de convenţii.

8.3.5. *Scale bazate pe raporturi şi coeficienţi*

Cel mai bun exemplu îl reprezintă măsura coeficientului intelectual QI, care se obţine prin compararea vârstei mintale cu vârsta cronologică. Specialiştii au demonstrat că vârsta mintală, luată izolat, nu dă un indice suficient de precis al dezvoltării mintale.

Pentru calcul se poate utiliza formula:

$$QI = \frac{\text{vârsta mintală}}{\text{vârsta cronologică}} \times 100.$$

Raportul se numeşte *coeficient de inteligenţă*. Denumirea va face carieră, fiind utilizată în varii domenii de activitate ştiinţifică şi practică.

QI-ul lui W. Stern, ca raport între vârsta mintală şi vârsta cronologică, nu trebuie confundat cu QI-ul lui Wechsler, care are, în general, valori mai înalte. Unui QI de 115, în varianta Wechsler, îi corespunde un clasament identic pentru subiecţi aparţinând unor grupe de vârstă diferite, pe când semnificaţia QI-ului lui W. Stern variază în funcţie de vârstă, deoarece abaterea standard a distribuţiilor acestor QI-uri nu este aceeaşi pentru toate vârstele.

În funcţie de problemele asociate abaterilor standard neechilibrate şi de faptul că dezvoltarea intelectuală nu are o relaţie liniară cu vârsta, raportul coeficientului creat de W. Stern nu mai este utilizat şi nu mai prezintă decât un interes istoric. În locul acestui

coeficient se utilizează scoruri standardizate bazate pe eşantioane reprezentative ale populaţiei de fiecare nivel. Aceste scoruri, numite şi *QI-uri deviate*, prezintă o medie de 100 şi o abatere standard de 15 (la scala Wechsler) şi de 16 (la scala Stanford-Binet), pentru fiecare categorie de vârstă.

8.3.6. *Scale bazate pe cuantumuri procentuale sau cuantile*

Cuantila de ordin p este acel element al unei serii ordonate de elemente care va depăşi un procentaj determinat, p, de elemente. Practic, este o „bornă" numerică separând două cuantumuri procentuale ale valorilor obţinute de un grup de referinţă la o anumită examinare prin intermediul unui instrument psihometric. Stabilirea sistemului de cuantile pentru un ansamblu de valori constituie o formă de etalonare. Ea permite situarea unui subiect, care a obţinut un scor determinat, într-o anumită clasă a subiecţilor pentru care testul a fost etalonat. Se va vorbi despre intercuantile (interdecile, de exemplu) pentru a desemna cele $n+1$ zone de distribuţie delimitate de cele n cuantile ale unui sistem de cuantile. Uneori se utilizează termenul de intercuantilă în loc de cuantilă. Distincţia clară se face prin precizarea că intercuantila este un interval numeric, pe când cuantila este reprezentată de o singură valoare numerică. Lucrările de specialitate ale lui Maurice Reuchlin (vezi în special 1976, pp. 67-68) consacrate statisticii aplicate în psihologie fac în mod clar această diferenţă de sensuri pentru cele două noţiuni.

Cuantilele şi intercuantilele numerotate cu 1 corespund, în general, în Franţa, celor mai bune rezultate, iar în SUA rezultatelor celor mai slabe. Dacă într-o distribuţie intercuantilele se vor reprezenta prin clase determinând intervale egale pe axa absciselor, se va asigura la această distribuţie grafică o formă rectangulară, efectivele teoretice ale tuturor intercuantilelor fiind egale prin definiţie. Este o aplicaţie a procedurii permiţând de a asigura o formă dată unei distribuţiei, modificând metrica variabilei măsurate.

La începutul procedurii de cuantilaj (formă particulară a etalonării) se alege o metrică, adică un număr de niveluri pe care scala permite să fie discriminate. Dacă o scală cuprinde n niveluri şi efectivul total al unui eşantion este N, efectivul fiecărui nivel va fi teoretic $\frac{N}{n}$. Scala în cuantile va corespunde, aşadar, unui model rectangular. Cele mai frecvente împărţiri în cuantile se fac între 1 şi 10 trepte (vezi tabelul 33).

Tabelul 34

Forme de cuantile

Număr de intercuantile	Număr de cuantile	Denumire tehnică
3	2	Trecilă
4	3	Quartilă
5	4	Quintilă
6	5	Sextilă
7	6	Septilă
8	7	Octilă
9	8	Nonilă
10	9	Decilă
100	99	Centilă

Dacă scala este constituită din 10 clase egale ca efectiv, acest efectiv va trebui să aibă 10% din numărul total de subiecţi. Se obţine astfel un decilaj. Intercuantila I corespunde primelor 10% din notele brute, cele mai slabe din distribuţie, intercuantila II conţine următoarele 10% din rezultate şi aşa mai departe, până la intercuantila X. Cuantila I va separa primii 10% dintre subiecţi de următorii 90%, cuantila II primii 20% dintre subiecţi de următorii 80% şi aşa mai departe, până la cuantila IX care va separa primii 90% de următorii 10%.

Avem ca exemplu situaţia în care un psiholog examinează un număr de 360 de subiecţi cu un anumit test, în vederea etalonării acestuia. Procedura se efectuează în faza construirii unui instrument de investigaţie psihologică sau în cazul reevaluării lui. Notele testului variază între 0 şi 35. Exemplul dat este unul imaginar şi nu va conţine primele cinci valori (1, 2, 3, 4, 5) şi ultimele trei (33, 34 şi 35). De obicei, în examinare se obţin foarte rar scorurile cele mai mici şi scorurile cele mai mari. Dacă datele ar fi fost reale, ar fi trebuit să se facă o serie de precizări cu privire la constituirea eşantionului de etalonare, adică la proveniența subiecţilor examinaţi, vârsta, nivelul de studii, sexul etc. De asemenea, ar fi trebuit să se facă precizări cu privire la tehnicile folosite pentru selectarea subiecţilor în eşantion. Înaintea operaţiunii de etalonare, datele se organizează crescător în variante de variaţie.

Tabelul întocmit în vederea stabilirii scalei de etaloane în 10 intercuantile (sau în 10 clase) va arăta astfel:

Tabelul 35

Determinarea cuantilelor (decile), pentru date organizate pe variante de variaţie

X	N	NC	NCT	Interdecila	Decila	Numerotare	% Interdecilă	Separaţie Decilă
6	3	3						
7	7	10						
8	4	14	36	0-11	11	I	10,88%	10,88%- -89,12%
9	6	20						
10	8	28						
11	11	39						
12	14	53						
13	17	70	72	12-13	13	II	8,61%	19,49%- -80,51%
14	16	86						29,21%-
15	19	105	108	14-15	15	III	9,72%	-70,79%
16	24	129						
17	22	151	144	16-17	17	IV	12,77%	41,98%- -58,02%
18	26	177	180	18	18	V	7,22%	49,20%- -50,80%
19	21	198						
20	24	222	216	19-20	20	VI	12,50%	61,70%- -38,30%

X	N	NC	NCT	Interdecila	Decila	Numerotare	% Interdecilă	Separaţie Decilă
21	18	240						
22	19	259	252	21-22	22	VII	10,27%	71,97%-28,03%
23	17	276						
24	14	290	288	23-24	24	VIII	8,61%	79,58%-20,42%
25	16	306						
26	14	320	324	25-26	26	IX	8,33%	87,91%-12,09%
27	11	331	331					
28	9	340	340					
29	7	347	347	27-32		X	12,09%	
30	6	353	353					
31	5	358	358					
32	2	360	360					

Notaţiile folosite în tabel sunt : X – nota brută, n – frecvenţa absolută a unei variante de variaţie, nc – frecvenţa cumulată absolută, nct – frecvenţa cumulată absolută necesară teoretic pentru calcularea intercuantilelor şi cuantilelor.

Jocul frecvenţelor cumulate absolute face să nu putem „tăia" distribuţia exact acolo unde trebuie. În acest caz, ne limităm în calcul la valorile cele mai apropiate de cele teoretice. Astfel, la prima decilă ar fi trebuit să ne oprim la valoarea 36. Valoarea 39 folosită de noi ca valoare delimitativă pentru prima decilă este cea mai apropiată de valoarea 36. Procedând în acest fel, am realizat un compromis metodologic. Acceptabilitatea compromisului poate fi stabilită prin testul χ^2 de ajustare.

Tabelul 36

Calculul statisticii χ^2

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
39	36	3	9	0,23
31	36	–5	25	0,81
35	36	–1	1	0,03
46	36	10	100	2,17
26	36	–10	100	3,85
45	36	9	81	1,80
37	36	1	1	0,03
31	36	–5	25	0,81
30	36	–6	36	1,20
40	36	4	16	0,40
			Total	11,32

Totalul reprezintă valoarea statisticii χ^2, după formula uzuală :

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}.$$

Coloanele tabelului indică desfăşurarea procedurii χ^2 de ajustare pentru obţinerea distanţei dintre efectivele teoretice necesare calculării decilelor şi cele observate practic. Coloana f_o desemnează efectivele observate, iar coloana f_t efectivele teoretice.

Pentru această valoare a lui χ^2 (11,32) găsim un $p > 0,10$, ceea ce duce la acceptarea ipotezei de nul, constatare care arată că distanţa dintre cele două serii de efective nu este semnificativă. Deci utilizarea efectivelor observate poate fi considerată un compromis acceptabil.

În cazul când dorim să stabilim o scală în centile, avem la îndemână o procedură puţin diferită. Pentru a pune în aplicare această procedură, trebuie să avem distribuţii ale căror amplitudini sunt de ordinul sutelor. Raţiunea construirii scalelor care fac apel la centile este de a asigura o discriminare foarte fină a subiecţilor, dar, în cazul când amplitudinea distribuţiei este sub 100, asigurăm o falsă fineţe a diferenţierii. Redăm, în continuare, un tabel ale cărui coloane indică paşii de urmat în procedura de calcul.

Considerăm un exemplu cu 137 de subiecţi care au fost examinaţi cu un anumit test :

Tabelul 37

Determinarea centilelor

Nr. crt.	X	f	f%	fc	fc%	Centila
1.	11	2	1,57	2	1,57	0,79
2.	12	5	3,94	7	5,51	3,54
3.	13	7	5,51	14	11,02	8,27
4.	14	11	8,68	25	19,69	15,35
5.	15	19	14,96	44	34,65	27,17
6.	16	26	20,47	70	55,12	44,89
7.	17	18	14,17	88	69,29	62,21
8.	18	15	11,81	103	81,10	74,50
9.	19	11	8,66	114	89,76	85,43
10.	20	7	5,51	121	95,28	92,52
11.	21	4	3,15	125	98,43	96,85
12.	22	2	1,57	127	100,0	99,22

unde X este scorul brut, f – frecvenţa absolută a variantelor de variaţie, f% – frecvenţa relativă a variantelor de variaţie, fc – frecvenţa cumulată ascendent, fc% – frecvenţa relativă cumulată ascendent, ultima coloană indicând percentila corespunzătoare.

S-a considerat un exemplu fictiv simplificat, deoarece în practică este foarte puţin probabil să se găsească un test la care să se poată obţine doar 12 valori distincte.

Procedura exemplificată de noi este folosită de autori reputaţi în domeniul psihometric ca David Magnusson (1975, p. 252), J.J. Bernier şi B. Pietrulewicz (1997, pp. 371-372). Prin această procedură, se încearcă să se amelioreze impreciziile metodei prezentate la calculul decilelor.

Coloana notată cu X conţine diferitele variante ale variabilei, obţinute în urma unei examinări cu un anumit test. A doua coloană, notată cu f, conţine frecvenţele diferitelor valori ale variabilei. În total, au fost cuprinse în analiză 127 de rezultate ale subiecţilor. Cu f% a fost notată frecvenţa simplă relativă, care se obţine prin înmulţirea cu 100 a rezultatului împărţirii fiecărei frecvenţe simple la 127 (numărul total de subiecţi examinaţi). Cu fc a fost notată frecvenţa absolută cumulată ascendent, iar cu fc% frecvenţa relativă cumulată ascendent. Valorile din această coloană se obţin prin înmulţirea cu 100 a rezultatului împărţirii fiecărei valori din coloana fc la 127.

Ultima coloană conţine transformarea în valori percentile a valorilor iniţiale ale testului. Cum se obţin aceste valori? Valoarea 13 este considerată mijlocul unui interval de clasă care se întinde de la 12,5 până la 13,5. Ca atare, valoarea percentilă este constituită din frecvenţa cumulată relativă anterioară valorii 13 şi din adăugarea a jumătate din frecvenţa relativă din dreptul valorii 13.

$$8{,}27 = 5{,}51 + \frac{5{,}51}{2},\ 15{,}48 = 11{,}2 + \frac{8{,}56}{2}\ \text{ş.a.m.d.}$$

Ce ne spune o valoare percentilă? Ea ne spune, de exemplu, că sub valoarea 16 se găsesc aproximativ 45% din totalul subiecţilor, iar peste această valoare se găsesc aproximativ 55% din totalul subiecţilor.

8.3.7. *Scale de scoruri standardizate*

Printre transformările curente pentru a face interpretabile scorurile brute, se numără şi următoarele:

- transformări liniare în scoruri standard, folosind media şi abaterea standard a distribuţiilor;
- normalizarea, constând în a transforma distribuţia scorurilor într-o distribuţie normală.

Practic, se divizează curba normală într-un număr de clase de lungimi egale, după care este uşor să se determine efectivele de subiecţi corespunzând fiecărei clase. Se pot întocmi tabele care să furnizeze procentajele teoretice de efective de clasă ale unei distribuţii normale în care se divizează ansamblul ordonat al valorilor în 7, 9, 11 etc. în părţi egale. Este suficient să se repartizeze măsurile originale într-un număr de clase, respectând procentajele. Aceste transformări sunt nonliniare.

Una dintre cele mai cunoscute transformări se referă la interpretarea măsurilor de interval, care poate fi facilitată efectuând o simplă transformare liniară de forma:

$$x_2 = ax_1 + b,$$

în care x_1 este măsura originală şi x_2 este măsura transformată.

$$a = \frac{\sigma_2}{\sigma_1} \text{ şi } b = m_2 - am_1,$$

m_1 şi s_1 fiind media şi, respectiv, abaterea standard a distribuţiei măsurilor originale, m_2 şi s_2 fiind valorile corespunzătoare pentru distribuţia dorită. În transformarea în scoruri centrate reduse (sau scoruri z), avem formula: $x_2 = (x_1 - m_1)/s_1$.

Astfel, măsurile pot face obiectul unei interpretări mai directe. Dacă aceste scoruri au o distribuţie apropiată de legea normală Gauss-Laplace, pentru a determina unde se situează rezultatul unui subiect în raport cu rezultatele subiecţilor unui grup de referinţă se aplică proprietăţile legii de repartiţie normală. De exemplu, în raport cu un subiect care va obţine un scor egal cu 2 într-o distribuţie normală de medie 0 şi o abatere standard egală cu 1, vom putea spune că „doar 2,3% dintre subiecţii grupului de referinţă obţin un scor superior celui obţinut de el".

În schimb, utilizarea scorurilor z în interpretarea unui scor observat are dezavantajul relativ că aceste scoruri pot avea valori negative, au o amplitudine foarte redusă şi se prezintă sub formă zecimală. Este motivul pentru care se utilizează, adesea, alte transformări liniare. Prezentăm, în continuare, o serie de transformări liniare posibile:

Tabelul 38

Transformări liniare ale scorurilor

Scor	m	σ	Corespondenţe între diferite scoruri transformate liniar					
Scor original	37,4	6,5	19	25	43	36	29	48
Scoruri z	0,00	1	– 2,83	– 1,91	0,86	– 0,22	– 1,29	1,63
Scoruri Stanford-Binet	100	16	54,71	64,48	113,78	96,56	79,32	126,09
Scoruri Wechsler	100	15	57,54	74,38	112,92	96,77	86,62	124,46
Scoruri H	50	14	10,37	23,29	62,06	46,98	31,91	72,83
Army General Classification Test	100	20	43,38	61,85	117,23	95,68	74,15	132,62
Scala Wechsler (pentru subteste)	10	3	1,51	4,28	12,58	9,35	6,12	14,89
College Board Scholar Aptitude	500	100	216,92	309,29	586,15	478,46	370,77	663,08
Scoruri Stanine	5	2	–0,66	1,18	6,72	4,57	2,42	8,76
Graduate Record Examination	500	100	216,92	309,29	586,15	478,46	370,77	663,08
Scoruri Stern	5,5	2	–0,16	1,68	7,22	5,07	2,92	8,76
California Entrance Examination Board	500	100	216,92	309,29	586,15	478,46	370,77	663,08
Scoruri T	50	10	21,89	30,92	58,82	47,85	37,08	66,31
% > Scor			99,76	97,20	19,50	59,70	90,15	5,26

Tabelul conţine exemple de transformări liniare aplicate la şase scoruri şi procentajul de subiecţi care au un scor superior scorului considerat, atunci când scorurile se distribuie normal. Exemplul este imaginat pentru o distribuţie care are o medie a scorurilor brute egală cu 37,4 şi o abatere standard egală cu 6,5. Coloanele m şi σ conţin (cu excepţia primelor valori) mediile şi abaterile standard pentru distribuţiile în care dorim să facem

conversia scorurilor. În practică, decizia este luată în funcţie de particularităţile situaţiei în care se află cercetătorul sau psihologul practician.

Explicităm, în continuare, câteva transformări uzuale.

Pentru obţinerea scorurilor z folosim formula:

$$z = \frac{X - m}{\sigma}.$$

Pentru o valoare brută egală cu 19 (prima exemplificare în tabel), vom avea:

$$z = \frac{19 - 37,4}{6,5} = -2,83.$$

Pentru obţinerea scorurilor T (introduse de McCall în 1922) vom folosi formula:

$$T = 50 + \frac{10}{\sigma}(X - m).$$

Pentru obţinerea scorurilor H (sau scoruri Hull), formula este:

$$H = 50 + \frac{14}{\sigma}(X - m).$$

Dacă facem transformarea z a unei distribuţii observate, care nu deviază semnificativ de la exigenţele curbei Gauss-Laplace, obţinem o nouă distribuţie, unde aproape toate valorile variază între – 3 şi + 3. Dacă transformarea se face în variabile T, noile date obţinute vor varia între 20 şi 80, iar atunci când conversia se face în variabile H se vor observa valori variind între 0 şi 100. Aceste note vor avea un plus de comoditate în utilizare. Optarea pentru un anumit tip de conversie a datelor înseamnă şi optarea pentru aceste limite de oscilaţie, indiferent ce amplitudini au distribuţiile notelor brute ale diverselor teste aplicate pe diverse eşantioane şi indiferent ce medii şi abateri standard utilizăm. Am subliniat mai sus condiţia ca aceste date brute să se distribuie aproximativ după o distribuţie normală. Denumirea frecventă a conversiilor bazate pe utilizarea mediei şi abaterii standard este de *scoruri standard*.

În obţinerea unor etaloane de bună calitate, intervin constrângeri de reprezentativitate şi de cost. Există o nevoie justificată de a le reactualiza periodic, date fiind schimbările temporare care se observă mai ales în randamentul la probele cognitive (autorii de specialitate vorbesc despre o accelerare biologică şi intelectuală a dezvoltării). După 1980, s-au produs teste psihologice computerizate care beneficiază prin program de o reetalonare permanentă, o dată cu introducerea unor noi date în memoria calculatorului. Scorurile standard prezintă avantajul de a menţine forma distribuţiei măsurilor originale. Din nefericire, dacă această distribuţie nu este normală, interpretarea scorurilor nu este în nici un fel facilitată prin transformare.

Un punctaj al unei distribuţii de valori brute, care a fost obţinut prin transformare liniară, are semnificaţie statistică clară. Poziţia relativă indicată de o anumită valoare standard a unei distribuţii de valori nenormalizate depinde în întregime de forma distribuţiei. Cu cât o distribuţie de scoruri se apropie mai mult de o distribuţie normală Gauss, cu atât mai sigur poate fi interpretată o valoare standard în distribuţie.

În scopul de a obţine o semnificaţie statistică clară a scorurilor individuale, se face adesea o normalizare a distribuţiei. Într-o distribuţie normală Gauss, se ştie exact câte procente de subiecţi există deasupra, respectiv dedesubtul unei valori pe o scală de o anumită medie şi o anumită unitate de măsură. Această calitate este exploatată în prezentarea

rezultatelor la testele psihologice în selecţie şi consiliere. Atenţia nu este concentrată pe comparaţia între diferite niveluri ale scalei, ci pe poziţia relativă pe care un individ o ocupă într-o distribuţie ale cărei proprietăţi sunt cunoscute.

O normalizare poate să fie realizată şi din alte motive. Când vrem să lucrăm cu diferite diferenţe inter- şi intraindividuale, trebuie să prezentăm poziţiile indivizilor în valori ale unei scale de interval. Nici o piedică insurmontabilă nu poate să apară la construcţia instrumentelor de măsură, dacă acestea trebuie să măsoare variabile de randament, care oferă de regulă distribuţii normale. La alte tipuri de variabile însă, este mai greu de realizat o distribuţie normală, de exemplu, la acele variabile măsurate cu ajutorul unui inventar de personalitate sau al unui chestionar de atitudini. Chiar când acceptăm premisele teoretice ale distribuţiei normale a variabilei cercetate pe un continuum ipotetic, obţinem, de regulă, distribuţii asimetrice.

Când acceptarea distribuţiei normale este îndreptăţită şi când se obţine o distribuţie asimetrică a valorilor brute, înseamnă că instrumentul de măsură nu indică îndreptăţit poziţia relativă. În astfel de cazuri, putem normaliza distribuţia asimetrică a valorilor obţinute, adică putem transforma forma distribuţiei, astfel încât să obţinem o distribuţie normală a valorilor.

În cazul în care avem distribuţii care deviază semnificativ de la distribuţia teoretică Gauss-Laplace, putem normaliza aceste distribuţii prin repartizarea variabilelor ordonate în limite impuse de procentele teoretice ale distribuţiei amintite anterior. Prezentăm un tabel cu rezultate normalizate:

Tabelul 39

Normalizarea datelor observate

Nr. crt.	Valoare brută	Valoare standard normalizată				
	X	F	FC	FCPM	P	Z
1	4	7	7	3,5	0,020	–2,05
2	5	4	11	9	0,051	–1,64
3	6	8	19	15	0,085	–1,37
4	7	6	25	22	0,125	–1,15
5	8	13	38	31,5	0,180	–0,91
6	9	16	54	46	0,262	–0,84
7	10	7	61	57,5	0,328	–0,45
8	11	11	72	66,5	0,380	–0,31
9	12	19	91	81,5	0,465	–0,09
10	13	13	104	97,5	0,557	0,14
11	14	24	128	116	0,662	0,42
12	15	14	142	135	0,771	0,61
13	16	6	148	145	0,828	0,95
14	17	16	164	156	0,891	1,23
15	18	7	171	167,5	0,957	1,72
16	19	4	175	173	0,988	2,25

Tabelul conţine, pe coloane, datele necesare pentru a normaliza o distribuţie. Semnificaţia coloanelor este următoarea: X – variabila brută distribuită pe variante de variaţie; f – frecvenţa simplă absolută; fc – frecvenţa cumulată absolută; fcpm – frecvenţa cumulată de punct mediu exprimată în valori absolute; P – proporţie obţinută prin împărţirea fiecărei valori din coloana fcpm la numărul total al subiecţilor (N=175); z – valoarea variabilei centrate reduse care separă într-o curbă normală Gauss proporţia respectivă de subiecţi de valoarea complementară până la 100 (de exemplu, valoarea z = –2,05 separă primii 2% dintre subiecţi de restul de 98%).

Indiferent de transformarea aleasă, trebuie să conştientizăm limitele acesteia. Cuantilajele au meritul simplităţii şi conduc la amplificarea diferenţelor minime între subiecţii din centrul distribuţiei (acolo unde se găsesc cei mai mulţi dintre subiecţi) şi la atenuarea diferenţelor la extremităţile distribuţiei (care nu cuprinde decât foarte puţini subiecţi). Normalizarea pare să posede numeroase avantaje: măsurile transformate sunt de nivel de interval şi se distribuie normal, ceea ce facilitează interpretarea şi aplicarea de metode de analiză care cer normalitatea distribuţiilor (analiză de varianţă, regresie multiplă).

Totuşi, nu trebuie să uităm că această metodă duce la schimbarea formei distribuţiei, fără să aibă întotdeauna raţiunea teoretică de a o face. Transformarea măsurilor originare permite ameliorarea interpretării ce poate fi făcută unui scor. Transformările normative sunt cele mai cunoscute şi cele mai vechi. Ele sunt legate de interesul pentru diferenţele interindividuale.

O imagine comparativă a scorurilor sub diverse transformări este următoarea:

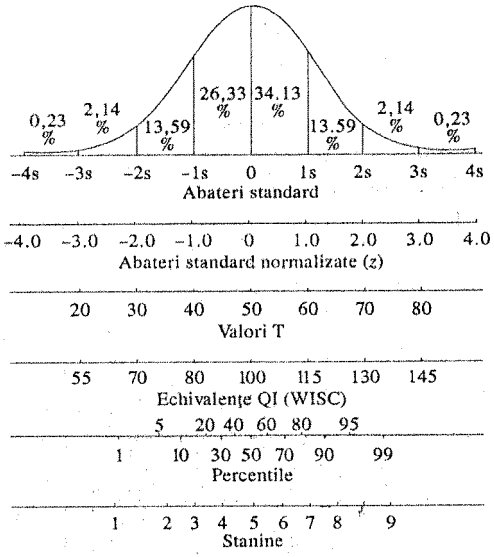


Figura 15. *Transformări normative de scoruri*

În practică, nu există transformări bune şi transformări rele. Diferitele metode aduc informaţii de natură diferită. Nevoile concrete ale psihologului practician sau ale cercetătorului sunt cele ce impun alegerea formei optime de transformare a notelor brute.

Capitolul IV

COMPUTERIZAREA TESTELOR PSIHOLOGICE

1. Generalităţi

Apariţia ştiinţei informaticii a constituit, ca domeniu al colectării, stocării şi prelucrării informaţiei, o revoluţie ştiinţifică şi tehnică având impact asupra tuturor activităţilor umane. La început, scepticismul a fost prezent în rândul unei largi pături de profesionişti, însă progresul informaticii a fost ferm şi impetuos, „măturând" din cale foarte multe „imposibilităţi". Raportul dintre informatică şi computer este acela dintre o disciplină ştiinţifică şi un produs tehnologic de înaltă performanţă care înglobează acumulările teoretice ale acesteia. Utilizarea computerelor în activităţi ştiinţifice, tehnice, administrative, economice devine un fenomen de masă şi a bulversat practic un mod de viaţă, restructurând un întreg univers de comportamente cotidiene. La această stare de fapt au contribuit mai multe faţete ale evoluţiei computerelor: ieftinirea echipamentului până la nivelul accesibilităţii de masă, mărirea capacităţii de memorare combinată cu accelerarea vitezei de prelucrare, extinderea manevrabilităţii acestora către grupuri de persoane lipsite de pregătire specializată în informatică, portabilitatea computerului şi, nu în ultimul rând, posibilitatea de a lucra în reţele din ce în ce mai cuprinzătoare. Firmele de hardware şi software au realizat mari progrese tehnologice şi economice pentru a face ca produsele lor să fie accesibile omului de rând şi a satisface cele mai neobişnuite aşteptări ale acestuia. Revoluţia informaţională propriu-zisă s-a realizat prin crearea microcomputerelor, care, organizate ulterior în reţele, au făcut posibile receptări informaţionale foarte prompte din cele mai îndepărtate regiuni ale globului.

Dezvoltarea informaticii după sfârşitul anilor '40 şi, mai ales, a microinformaticii, începând cu anii '70, a avut efecte sensibile asupra aplicării testelor psihologice. Într-o primă fază, când nu se dispunea decât de mari sisteme de gestiune foarte puţin prietenoase, informatica a fost utilizată ca un mijloc de calcul economic, permiţând dezvoltarea metodelor de analiză a datelor, analiză factorială mai ales, dificil de practicat în absenţa sa. Informatica a fost utilizată în acel timp mai ales pentru a corecta testele. Asocierea informaticii şi a sistemelor de lectură optică s-a făcut în 1955 şi a permis o mare economie de timp în tratamentul testelor colective. Aceste prime aplicaţii nu merg dincolo de automatizarea procedurilor manuale anterioare. Posibilităţile de a calcula scorurile diverse sau de raportare la etalonări variate sunt puţin utilizate. Progresiv, procedurile informaticii încep să fie exploatate.

Situaţia de administrare este adesea informatizată. Subiectul răspunde la întrebări derulate pe un ecran, manipulând o claviatură, un mouse sau un creion optic. Se pot propune şi situaţii mai complexe decât acelea ale testelor clasice (de exemplu, situaţii care simulează activităţile profesionale). Dispozitivele informatice permit, de asemenea, înregistrarea parametrilor activităţii mintale : numărul şi natura erorilor, timpii de reacţie la prezentarea stimulilor, durata rezolvării, ordinea operaţiilor efectuate etc. Aceste informaţii permit să se reconstruiască, cel puţin parţial, strategiile utilizate de către subiecţi şi adesea să se rafineze diagnosticul. Marile capacităţi ale memoriei ordinatorului permit stocarea informaţiilor primite în cursul administrării şi, astfel, constituirea unor bănci de date (cu actualizarea permanentă a etalonării). Se dezvoltă, din ce în ce mai susţinut, asistarea automată la interpretare, rezultatele subiectului sunt comentate şi se furnizează elemente susceptibile de a le interpreta. Interpretările propuse pot fi fondate pe calcule statistice (coeficienţi de validitate empirică, de exemplu) sau pe o simulare a activităţii psihologului (activitatea de consiliere în cadrul unei practici clinice).

Diversele posibilităţi oferite de informatică sunt încă larg subutilizate şi multe dintre realizări nu au depăşit stadiul experimental. Este probabil ca în viitor, conjugată cu aportul psihologiei cognitive, tehnologia informatică să reînnoiască în profunzime testele de inteligenţă.

2. Contribuţii aduse computerizării testelor psihologice

Computerul a devenit o prezenţă obişnuită, aproape indispensabilă în laboratoarele de psihologie. Ameliorarea gestiunii experimentale este o evidenţă care nu mai trebuie demonstrată. În psihologie, începuturile au fost făcute în domeniul prelevării, corectării şi evaluării rapide a datelor de la testele şi chestionarele psihologice, cu mai bine de trei decenii în urmă. De asemenea, programele statistice au constituit una dintre primele aplicaţii în psihologie. Limitarea posibilităţilor de aplicare s-a făcut în funcţie de nivelul tehnologic al constructorilor de computere. Lărgirea posibilităţilor a dus la conducerea prin software specializat a cercetărilor din domeniul psihologiei cognitive şi al celei fiziologice, în dezvoltarea procedeelor de testare psihologică şi simulările pe computer. Cele mai multe aplicaţii s-au efectuat totuşi în diagnosticul asistat de calculator, care este sinonim cu dezvoltarea şi introducerea testelor computerizate şi a sistemelor de interpretare a rezultatelor. În cadrul acestuia, se disting aspectul de corectare a testelor şi de transformare corespunzătoare a datelor brute, efectuarea propriu-zisă a probei şi redarea rezultatelor. Multiplicarea posibilităţilor s-a accelerat o dată cu înlocuirea unor instalaţii mari cu computerele personale. Minnesota Multiphasic Personality Inventory (MMPI) a fost unul dintre primele instrumente metodologice beneficiare ale unor astfel de tratamente informatice.

Prelucrarea automatizată a datelor în domeniul psihodiagnosticului a început în anii '20 prin prelucrarea datelor oferite de aplicarea inventarului SVIB (Strong Vocational Interest Blank) cu ajutorul maşinilor de calcul mecanice. În 1946, Elmur Hankes a construit un computer analogic pentru prelucrarea automată a SVIB, care presupunea calcularea scorurilor şi stabilirea de profiluri. Tot în acel an, se produce evaluarea automată

a MMPI, cu IBM 805, experienţă care se dovedeşte deosebit de economicoasă în timp şi bani. Professional Examination Division, care ţine de Psychological Corporation din New York, creează primul *Machine Scoring Service*. Despre acest eveniment scriu, în revistele americane de psihologie clinică, C.E. Davis (1947), L.A. Drake (1947), W.G. Dahlstrom şi G.S. Welsh (1960) (*cf.* Spreen, 1963, p. 18). În primii ani ai deceniului şase, combinarea scannerului optic şi a computerului au dus la o prelucrare fără greşeli a foilor de răspuns oferite de MMPI. Primul sistem automatizat în sens modern a fost introdus în clinica Mayo din Rochester, Minnesota. Itemii chestionarului MMPI erau prezentaţi pe cartele de răspuns speciale, prelucrate de subiecţi şi apoi introduse printr-un scanner în computerul central. Un program de evaluare elabora valori normalizate pentru 14 scale MMPI şi comenta valorile scalelor prin ordonarea afirmaţiilor verbale dintr-o bibliotecă-program ce cuprindea 62 de specificări, dintre care majoritatea au fost asociate cu cotele de pe scalele MMPI. Câteva dintre aceste afirmaţii cuprindeau şi verdicte configurale. Acestea se referă la interpretări bazate pe *pattern*-uri specifice ale scorurilor scalei. Sistemul era utilizat în principal în cadrul cercetărilor medicale primare, ca procedeu *screening*. Scopul introducerii era o identificare rapidă şi puţin costisitoare a pacienţilor care trebuiau să fie îngrijiţi, respectiv cercetaţi psihologic în continuare. Numărul pacienţilor examinaţi cu acest sistem a fost de ordinul miilor. Succesul sistemului Mayo de la începutul anilor '60 a servit drept impuls pentru computerizarea altor probe. Astfel, Piotrowski a creat în anul 1964 un sistem computerizat pentru interpretarea testului Rorschach. O listă de 325 de parametri cuprindea codificarea răspunsurilor individuale. Interpretarea a avut drept fundament aceşti parametri, nu răspunsurile brute.

Un al doilea sistem, elaborat în ordine cronologică, a fost aşa-numitul Fowler-System, elaborat de R.D. Fowler în 1968-1969, la University of Alabama, cu scopul de a simula procesul de decizie în domeniul interpretării testelor MMPI şi de a produce, pe baza rezultatelor individuale şi a configuraţiilor anumitor scale, o expertiză care ar fi trebuit să se apropie de diagnosticul clinic în conţinut şi stil. În cadrul unui centru de servicii comerciale, au fost realizate în 17 ani un total de peste 1,5 milioane expertize cu ajutorul sistemului Fowler (folosit de fiecare al patrulea psihiatru şi psiholog din Statele Unite).

Între 1966 şi 1969 Finley-System a fost dezvoltat la University of Kentucky. J.C. Finley şi colaboratorii au construit scale speciale suplimentare, care au fost introduse în evaluare. Sistemul producea interpretări (pe baza configuraţiilor scalelor MMPI) ale configuraţiilor scalelor suplimentare, ca şi pe baza valorilor supraordonate, dobândite prin intermediul analizei factoriale. În interpretare, se fac comparaţii cu CPI (California Psychological Inventory), care fusese administrat anterior. Pentru acest sistem a fost realizată o normare. După fiecare cerere, sunt create până la 19 feluri de expertize, al căror conţinut însumează până la 16 pagini. Sistemul a fost comercializat.

Caldwell-System (A.B. Caldwell, 1971) produce interpretări pentru configuraţiile specifice asupra scalelor MMPI originale şi asupra câtorva scale suplimentare. Raportul asupra rezultatelor este diferenţiat în funcţie de diferite forme de fundamentare (categorii psihodinamice, descrieri de comportamente, simptome) şi conţine o consiliere despre indicaţiile de tratament posibile. Şi acest sistem a fost comercializat.

În 1974, apare Lachar-System, care pune la dispoziţie o cuprinzătoare bibliotecă de prescripţii interpretative şi de reguli de evaluare şi decizie pentru producerea expertizelor verbale, în special pe baza ipotezelor diagnostice intuitive. Evaluarea comercială a fost realizată prin Automated Psychological Service, ulterior prin Western Psychological Services.

La sfârşitul lui 1970, a fost dezvoltat, la University of Minnesota, Butcher System (comercializată prin National Computer Systems). Sistemul foloseşte o cantitate suplimentară de scale, în parte dezvoltate pe baza cunoştinţelor empirice, precum şi indici privind datele demografice ale subiecţilor. Expertiza ia forma unei validităţi de profil, de clusteri de simptome şi de stabilitate comportamentală, conţinând propuneri pentru decizie diagnostică şi pentru posibilităţi de tratament. Sistemul poate elabora variante de expertiză ce pot fi cerute de diferite instituţii, în funcţie de specificul unei anumite probleme. Avantajul sistemului – şi cauza folosirii lui pe scară largă – a constat în evaluarea automată a testului, presupunând costuri relativ mici pentru client. Trimiterea chestionarelor MMPI la un centru comercial de testare pentru interpretare automatizată a datelor poate reduce în mod considerabil activitatea personalului auxiliar din clinici, activitate care cere timp şi efort. Doar diagnosticianul clinic era în situaţia de a lua în considerare ansamblul datelor şi rezultatelor existente la fiecare caz de diagnosticat, în toată complexitatea sa. În acest sens, sistemele automate pot contribui la sporirea validităţii deciziei diagnostice. Un alt mare avantaj constă în aceea că evaluările automate sunt mai puţin accidentale faţă de erorile de măsură nesistematice ce pot fi produse de către corectorii umani, datorită neatenţiei şi oboselii accentuate.

După anii '70, psihologii au realizat că ordinatoarele pot fi integrate în toate etapele procesului de evaluare psihologică. S-a observat deja că avantajele pe care le oferă calculatorul sunt numeroase, mai ales în efectuarea automatizată a testului (precizia măsurării, evidenţierea unor noi indicatori comportamentali, micşorarea timpului de administrare, simplificarea activităţilor de rutină pentru psiholog, fiabilitatea concluziilor). Realizările de până atunci au scos în evidenţă faptul că evoluţia raporturilor între psihometrie şi informatică depinde de doi factori, şi anume evoluţia tehnologică, pe de o parte, şi evoluţia teoriilor şi concepţiilor psihologice asupra măsurii, pe de altă parte. În anii '70, se conturează majoritatea cerinţelor informaţionale, care sunt rezumate de Jean Luc Roulin (1990, p. 266) :

- un nou sistem de prezentare a informaţiei (cel mai adesea redus la un ecran de vizualizare) ;
- un sistem de înregistrare a răspunsurilor multiple (claviatură, atingerea ecranului), lărgind posibilităţile testelor creion-hârtie ;
- un sistem rapid de tratament al informaţiei ;
- un sistem de analiză a datelor ;
- un sistem de stocare a informaţiilor.

Sunt conturate trei căi de dezvoltare în automatizarea testelor :

a) reluarea probelor clasice la care se adaugă noi indicatori ;
b) crearea de noi indicatori comportamentali prin intermediul unor probe originale pentru evaluarea diferenţelor inter- şi intraindividuale ;
c) efectuarea adaptată automatizat.

În 1975, Williams şi Johnson au descris utilizarea unui computer principal (*mainframe*) care era legat cu câteva terminale pentru evaluarea a 17 pacienţi pe zi. În mod obişnuit, pacienţii completau MMPI, BDI (Beck Depression Inventory), teste de memorie, de inteligenţă, anamneză etc. Autorii amintiţi au demonstrat că rapoartele computerizate au fost generate în jumătate de timp şi cu jumătate din cheltuieli.

În anii '80, CAPA (Computer Assisted Psychological Assessment) a devenit prevalentă, astfel încât, practic, orice test psihologic putea fi interpretat cu ajutorul unui calculator electronic.

În 1985, R.D. Fowley aduce în atenţie o serie de aspecte problematice pentru decizia introducerii sistemelor automatizate :

- o posibilă dezumanizare a situaţiei de examinare psihologică ;
- posibile probleme etice şi juridice legate de utilizarea serviciilor comerciale ;
- posibile probleme de validare, referitoare la interpretarea valorii testelor şi calitatea expertizelor produse.

Problema utilizării de către neprofesionişti a rapoartelor computerizate prezintă într-adevăr un pericol, prin situaţia de supraapreciere a obiectivităţii datorate calculatorului. Mai ales interpretările urmează reguli precise pe care creatorii lor nu le publică, invocând secretul profesional şi protejarea dreptului de autor. O altă situaţie nefirească este că regulile de decizie nu fac obiectul unor reevaluări periodice, în funcţie de evoluţia activităţilor de validare.

După opinia lui Fowley, în examinarea psihologică se decide mai ales în funcţie de felul contactului personal şi de situaţia de consiliere legată de aceasta. Acordând o mare atenţie modului cum este structurată uman această consiliere, autorul amintit recomandă mai puţin realizarea şi evaluarea computerizată a testului. Totuşi, ipoteza unei posibile dezumanizări a situaţiei de cercetare diagnostică prin introducerea procedeelor computerizate pierde teren datorită atitudinii de acceptare, chiar preferare din partea subiecţilor, precum şi accentuatei motivări şi sincerităţii persoanelor cercetate, în comparaţie cu procedeele tradiţionale. Fowley mărturiseşte că problema introducerii testelor computerizate se pune mai degrabă pentru diagnosticieni, decât pentru subiecţi.

Computerizarea în domeniul psihodiagnosticului nu se aplică doar la chestionarele de personalitate, ci şi la o serie de teste aparative de randament psihic. Noţiunea de *test aparativ* este puţin uzitată în literatura românească de specialitate, dar are răspândire largă în literatura germană şi anglo-saxonă. În mod concret, este vorba despre instalaţii de examinare, care furnizează stimulii şi înregistrează rezultatele examinării, gama acestora fiind deosebit de diversificată.

Până la mijlocul anilor '50, se producea aparatură construită exclusiv pe principii electromecanice. După această dată, apar generaţiile de aparate electronice convenţionale şi, după anul 1980, se introduc progresiv aparate de examinare psihologică folosind microprocesoarele. Puterea de memorare a generaţiilor succesive de calculatoare, precum şi viteza de lucru au influenţat decisiv amploarea aplicaţiilor în domeniul testului psihologic. Se poate vorbi, în domeniul construcţiei testului psihologic de randament, despre o tendinţă constructivă extensivă şi despre una intensivă. Prima are în vedere mai mult diversificarea modurilor de stimulare psihologică şi mărirea indicatorilor de măsură memoraţi şi prelucraţi de calculator. Executarea pe acelaşi calculator a diferite programe de investigaţie psihologică şi integrarea calculatorului în procesul de colectare a unei diversităţi de informaţii presupuse relevant psihodiagnostice fac parte din această primă tendinţă. Astfel, calculatorul poate integra şi informaţii de natură psihofiziologică şi anamnestică. Cea de-a doua tendinţă are în vedere, cu prioritate, punerea în aplicaţie a unor noi concepţii psihodiagnostice, adăugarea de indicatori neutilizaţi până atunci, rafinarea tehnicilor de analiză a datelor şi obţinere a rezultatelor. Creaţiile care reunesc cele două tendinţe în chip armonios sunt cele mai dezirabile în câmpul psihologiei aplicate.

Aspecte de un deosebit interes sunt relevate de Garnot Schufried, într-un articol (1987, pp. 195-199) referitor la noile tendinţe în dezvoltarea aparatelor având ca destinaţie examinarea psihologică. Autorul semnalează că, în construcţia aparatelor amintite mai sus, a apărut o nouă fază interesantă, cea a reorientării funcţionale, şi indică următoarele avantaje aduse de utilizarea aparatelor:

- comasarea diferitelor procedee de examinare prin intermediul unui singur aparat;
- posibilitatea interacţiunii între subiect şi aparat;
- eliminarea efectelor datorate psihologului examinator;
- automatizarea instructajului subiectului, care cuprinde atât efectuarea, cât şi evaluarea;
- costurile relativ mici ale computerelor.

Întregul spectru de calculatoare (PC, ZX 81, Commodore, IBM/PC şi compatibile cu acesta, PDP 81) din anii '80 au un bun raport randament – preţ. În schimb, pot apărea diferite funcţionări deficitare. G. Hopmann şi N. Gallery au întreprins studii în cadrul cărora se analizează o configuraţie pentru măsurarea timpului de reacţie în funcţie de locul apariţiei stimulului. În diferite locuri pe ecran apar aleatoriu diferite cifre, la care subiectul trebuie să acţioneze pe o anumită tastă. Pentru evidenţierea unei diferenţe de timp de reacţie care interesează (stabilit sub o milisecundă), a fost creată o rutină pentru măsurat timpul, cu pasul de 12 ms.

Analiza exactă a programului arată că durata emiterii unui semnal presupune o întârziere datorată caracteristicii hardware. Durata exactă a întârzierii depinde de funcţionarea calculatorului, de program şi de sistemul de operare. Oscilaţiile se fac simţite la nivelul milisecundelor pe unul şi acelaşi calculator. Nu s-a luat în seamă faptul că este nevoie de 20 de milisecunde pentru rescrierea ecranului. Sunt deci importante cunoştinţele şi cuprinzătoarea înţelegere tehnică privind funcţionarea sistemului hardware şi software, necesare în scopul unei construcţii lipsite de artefacte.

La aparatura existentă în acel timp (1987), următoarele aspecte erau considerate ca fiind deficitare:

- tastatura nu era indicată ca mediu adecvat de reacţie;
- frecvenţa de *refresh* a ecranului;
- configurarea nesatisfăcătoare a ecranului;
- încetinirea vitezei sistemului, mai ales când se utilizează itemi grafici;
- măsurarea timpului de rezolvare sub 1/20 secunde era problematică;
- întârzieri şi blocaje survenite la înregistrarea indicatorilor externi de reacţie.

Fireşte că, în funcţie de scopul aplicării, pot fi considerate nesemnificative unele sau altele dintre punctele marcate ca deficitare.

Garnot Schufried evidenţiază anumite tendinţe de corectare a unor astfel de defecte existente la unele firme constructoare americane, care dotează calculatoarele cu un panou cu tastele „da-nu", „nu ştiu" etc. Atunci când pe calculator sunt aplicate predominant chestionare din domeniul personalităţii, subiecţii au avut o ordonare a tastelor în funcţie de întrebările de pe ecran. De cele mai multe ori, rezultatele au fost comparate direct cu normele şi adesea au fost verbalizate. În unele institute şi clinici din RFG, existau mici calculatoare prevăzute cu tastatură şi indicatoare luminoase. Aceste disponibilităţi individuale indicau clar că trebuie realizat un compromis între o dispunere experimentală optimală şi posibilităţile tehnice.

Soluţia la această situaţie nemulţumitoare o reprezintă conceperea unui sistem computer înzestrat cu un hardware special şi cu sistem special de funcţionare, care să fie adaptat optimal la exigenţele psihologice. La acea vreme aproximativ 400 de astfel de instalaţii erau introduse în practică. Dezavantajul consta în faptul că preţurile acestor sisteme erau foarte mari. Construcţiile speciale pun clar problema compatibilităţii cu proba clasică.

ART-90 constituia, în concepţia lui Schufried, cea mai consecventă dezvoltare a acestui concept. Aparatul a fost dezvoltat în Germania de *Curatorium*-ul de siguranţă a circulaţiei în colaborare cu o firmă austriacă. ART-90 este o probă compozită ce reuneşte un pupitru-test structurat ergonomic, monitor pentru candidaţi, „cognitron" (nu se fac precizări în privinţa acestei denumiri), aparat de măsurat reacţiile, aparat de determinare, tahistoscop, magnetofon şi un aparat pentru percepţia periferică.

Un program-test bun este compus, în concepţia autorului, din mai multe secvenţe. Faza de instrucţie este concepută în manieră progresivă după principiul învăţării programate. Doar cei care au înţeles exigenţele de solicitare pot să ajungă să parcurgă faza testului. Programul trebuie să memoreze timpii de rezolvare a itemilor individuali.

Programul de evaluare, care este realizat ulterior, exprimă rezultatul în note brute, iar când preexistă norme, realizează comparaţii. Pentru câteva programe sunt verbalizate rezultatele şi tipărite interpretările.

Programul de utilizare permite administrarea unei întregi baterii de teste la care dificultatea testului poate fi variată printr-un parametru. De o deosebită însemnătate practică sunt programele statistice efectuate pe datele acumulate.

Un rezultat al colaborării dintre psihologii şi tehnicienii Academiei de Medicină din Magdeburg (secţia de biomedicină tehnică şi clinica pentru neurologie şi psihiatrie) este funcţionarea, din 1983, a unui sistem de aparate asistate computerizat pentru psihodiagnostic.

Complexul aparativ era intitulat Combi-2 şi era instalat pe computere personale având sistemul de operare SCP (compatibil CP/M) (vezi în acest sens Weber, Leo şi Regel, 1987, pp. 200-206). Combi-2 funcţionează ca aparat de determinare şi îşi propune următoarele scopuri:

- cercetarea coordonării vizual-motorii;
- realizarea unui test tremometru;
- detectarea tempoului psihomotrice (*tapping*);
- precizarea fuziometriei;
- realizarea ecranului de protecţie pentru cuprinderea motricii vizuale, aptitudinilor de învăţare şi memorare a comportamentului complex de reacţie;
- realizarea de chestionare pe calculator.

Componenta hardware principală a sistemului este pupitrul stimul-reacţie cu posibilităţi de aplicare a stimulilor.

Reacţiile sunt posibile prin intermediul a cinci taste robuste, cu suprafaţă adecvată ergonomic, cu o cursă scurtă, care se găsesc într-o ordine definită, tasta *tremor* şi o tastă *tapping*. Posibilităţile de informare pentru subiecţi se produc prin intermediul a patru câmpuri de informaţie luminoase scrise (de exemplu, fals, corect, mai repede), iar stresările şi iritările se produc prin zgomote pe un magnetofon.

Sistemul cuprinde 58 de sarcini de determinare pentru următoarele aspecte:

- măsurarea timpului de reacţie şi a siguranţei de reacţie, la reacţii cu alegeri simple şi multiple, la stimulii optici sau acustici;

- observarea aptitudinilor de concentrare;
- verificarea capacității de învățare.

Prin realizarea colectării multiple de elemente, au fost dezvoltate forme paralele pentru un număr de sarcini. Programul de determinare permite următoarele variante de aplicare:

a) reacția următoare este declanșată când este apăsată tasta centrală 1. Este posibilă o diferențiere între timpul de mișcare și cel de inițiere;
b) toate tastele sunt la fel de indicate. Subiectul precizează, ca la punctul a), rapiditatea testului. Trebuie să se răspundă la fiecare reacție;
c) calculatorul precizează tactul personal. Reacțiile sunt produse la distanțe temporale prestabilite constant;
d) calculatorul precizează tactul adaptativ. Începând cu o distanță de T=3s între prima și a doua reacție, aceasta este mărită sau micșorată procentual, după fiecare reacție falsă sau corectă. Distanța T a următoarei reacții devine astfel F×T;
e) genul special de program pentru precizarea aptitudinii de învățare.

Variabilele prioritare în dialogul cu calculatorul sunt următoarele:

- numărul de reacții;
- durata reacției în ms;
- distanța temporală între reacție și următoarea reacție (constant și pseudostohastic);
- evidențierea erorilor în câmpul informațional.

Programele permit analiza foarte amănunțită a desfășurării testului pe diferite fațete. Sunt prezentate valorile numerice, mediile, împrăștierile tuturor timpilor măsurați, cota-parte de reacții false, precum și graficele corespunzătoare.

Pe Combi-2 au fost implementate următoarele chestionare:

- FPI (*Freiburger Personlichkeitsinventar*);
- PPKV (*Psychopathologisches Kurzverfahren*);
- VNPI (chestionar pentru investigarea labilității vegetative, a nevrotismului și a altor caracteristici ale personalității).

Sistemul poate fi extins. Pe lângă prelucrarea algoritmului de evaluare corespunzător fiecărui test, este posibilă o cercetare a răspunsurilor într-un sistem de evaluare complex.

Dintre testele administrate de Combi-2, exemplificăm două:

- Pentru *Pursuit Test* este necesară montarea unei interfețe video pentru monitorul pacientului și a unei manete de ghidare (interfață în pupitru excitație-reacție). Pe monitor apare desenat un pătrat, care se mișcă într-o direcție, pseudostohastic, pe suprafața ecranului. Sarcina subiectului este de a conduce un pătrat mai mic, deplasat prin maneta de ghidare, în interiorul primului pătrat al testului. Se comite o eroare atunci când pătratul mai mic se găsește în afara pătratului mai mare. Variabilele de control sunt mărimea, rapiditatea de deplasare și mișcarea pătratului mai mare pe ecran. Rezultatele sunt exprimate numeric și grafic în secvențe de 30 de secunde: rapiditatea, numărul de erori și timpul în care se găsea punctul în afara pătratului testului în procente la 30 de secunde (*out time*). Ca modalitate specială, este posibilă realizarea unui test adaptativ. Rapiditatea este variabilă, în acest caz, în funcție de prezența peste o cotă-parte procentuală în afara spațiului pătratului mare. *Pursuit test* permite afirmații diagnostice pentru coordonarea vizual-motorie.

- Testul *labirint* permite testarea a 12 imagini-labirint, cu grad de dificultate diferit, în patru variante paralele. Labirintul apare pe ecran și subiectul are posibilitatea de a memora configurația lui în timpul de prezentare la alegere. După ce dispare labirintul, iar pe monitorul pacientului apare un grilaj care marchează părțile închise, respectiv deschise, ale labirintului, își face apariția un punct în colțul din stânga jos, punct ce trebuie adus cu ajutorul manetei de ghidare în colțul din dreapta sus. Dacă se depășește linia de delimitare a unui drum închis, se contorizează o eroare și se măsoară perioada de timp în domeniul interzis. Se consideră eroare și când punctul se găsește din nou în pătratul în care s-a produs o eroare. O variantă de program constă în prezentarea unui șir de puncte care dispar când subiectul face o eroare. Cu aceasta candidatul primește o informație vizuală suplimentară în privința drumului urmat în efectuarea testului.

Programul Combi-2 are câteva rutine care permit schimbările de program prin comandă menu, ca și identificarea subiectului. Sunt permise manipularea și evaluarea statistică a băncilor de date cu software comercial dBASE.

O contribuție demnă de semnalat aparține lui Mario Caruso (1987), cercetător din secția de psihologie a Universității „Karl Marx" din Leipzig. El se referă la produsele software din domeniul psihodiagnosticului, utilizate în fosta RDG, o dată cu apariția microcomputerelor, cu utilizare principală în psihologia experimentală și în facilitarea activității psihologilor practicieni.

Procedeele sprijinite computerizat pentru psihodiagnostic și terapie găsesc o acțiune largă când se realizează un consens în privința următoarelor probleme:

1) alegerea configurației de aparate și a instalațiilor periferice suplimentare;
2) compatibilitatea și folosirea multiplă a componentelor software;
3) elaborarea unor produse software aplicabile cu interfață intuitivă;
4) normarea testelor sprijinite computerizat și precizarea efectivității la programele de antrenament sprijinite computerizat;
5) posibilitatea protecției împotriva copierii.

Mario Caruso indică existența, în domeniul internațional și național, a unei întregi serii de configurații de aparate, care au fost dezvoltate pentru psihodiagnostic și au intrat în exploatare. După autorul amintit, în anul 1987, în interiorul fostei RDG erau dezvoltate și aplicate produse psihodiagnostice sprijinite computerizat, psihologii lucrând la 60 de dispozitive, dezvoltate pentru șapte configurații fundamentale diverse:

- Combitest 2 (Krause, 1986)[1]
- K 1520
- Microcomputer Kc 85/1-3 (Hub, KMU);
- Wiener Testsystem de la firma Schufried (Grubh);
- Mc 80 Basis (VMD Halle);
- Z 1001

Aceste configurații de bază, cu diverse aparate periferice, au o dezvoltare autonomă. Autorul intuiește viitoarele aplicații datorate apariției unor sisteme computerizate mai performante în privința memoriei și vitezei de prelucrare a informației.

1. Prezentăm între paranteze numele autorilor sau al firmelor producătoare.

Problema pe care şi-o pune Caruso este aceea a compatibilităţii rezultatelor în cazul rulării aceloraşi programe pe instalaţii hardware diferite, aşadar problema influenţei speciale a componentelor hardware asupra rezultatelor testului, precum şi influenţa efectelor de antrenament cu aparatura. Autorul apreciază (bazându-se pe informaţii provenind de la H. Cooper, G. Stiller, H.J. Jackel, din anul 1982) că între 80% şi 85% din costurile introducerii computerului în examinare revin produselor software. Date fiind costurile, se impune exigenţa unei configuraţii unitare a aparaturii.

Exigenţele acestei configuraţii au fost reliefate din experienţa grupei de cercetare „Diagnostic computerizat" a KMU (firmă specializată). Una dintre ele susţine că în viitorul previzibil se vor construi sisteme adaptate doar nevoilor psihologilor. Alte exigenţe sunt:

- dezvoltarea produselor software trebuie să aibă mai multe direcţii; introducerea relativ autonomă a unor componente de program pentru cei care au un microcomputer cu anumite periferice;
- organizarea feedback-ului în testele adaptative;
- evaluarea temporală a rezultatelor, fără a fi necesară memorarea rezultatelor (pentru seturi de date strict delimitate, utilizarea unor procedee statistice);
- fiecare psiholog practician să fie pus în situaţia de a putea selecta procedeul computerizat.

Cu toate că realizarea exigenţelor amintite, cu înalte calităţi, contribuie mult la răspândirea psihodiagnosticului computerizat, acestea vor cauza şi o clară osificare a psihodiagnosticului practic, când prin programe nu sunt convertite concepte mai greu aplicabile fără calculator. Aplicarea acestor concepte înseamnă realizarea de teste adaptative, aplicarea de paradigme ale psihologiei experimentale în psihodiagnostic, care evaluează simultan reacţiile subiectului în procesul psihodiagnosticului.

Normarea testelor şi examinarea efectivităţii programelor de *training* sprijinite computerizat constituie alte două exigenţe semnalate de Mario Caruso. Partea de reducere a efortului este evidentă. În cazul în care testele creion-hârtie au fost transpuse computerizat, nu există nici un fundament de a accepta normele manuale creion-hârtie pentru computerizare şi se impune realizarea unei noi normări.

O problemă considerată serioasă este nevoia protecţiei contra copierii programelor psihodiagnostice şi răspândirii lor în rândul neiniţiaţilor.

3. Echivalenţa versiunilor computerizate

În privinţa echivalenţei versiunilor computerizate şi tradiţionale ale MMPI, s-au publicat un număr apreciabil de cercetări empirice, cu rezultate şi concluzii diferite[1].

Gabrielle Franke (1997, pp. 332-356) prezintă rezultatele a două studii experimentale în privinţa practicii prezentării asistate computerizat a inventarelor psihodiagnostice.

1. Vezi Jürgen Hagebock, *Computerunterstutzten Testen*, Hogrefe-Verlag für Psychologie, Göttingen, 1994. În această carte sunt prezentate contribuţiile lui Y.S. Ben-Porath şi J.N. Butcher, 1986; M.J. Burke şi J. Normand, 1987; Butcher, Keller şi Bacon, 1985; P.J. Hofer şi B.F. Green, 1985; K.L. Moreland, 1985, 1987; R.H. Rosensky, L.F. Honor, K. Rasinsky, S.M. Tovian şi G.I. Herz, 1986; L.M. Honaker, 1980.

Autoarea germană face o trecere în revistă a problematicii probelor computerizate, din care trage concluzia existenţei a două tipuri de echivalenţe: una psihometrică, exprimată prin intermediul valorilor cantitative, şi alta experienţială, exprimată prin ansamblul trăirilor specifice stării de examen. Rezultatele din literatura de specialitate sunt contradictorii în privinţa diferitelor categorii de teste şi a criteriilor de apreciere psihodiagnostică, dar majoritatea studiilor de referinţă indică preferinţa accentuată a subiecţilor pentru versiunile computerizate ale testelor psihologice. Cercetarea lui Gabrielle Franke s-a realizat pe un număr de 600 de subiecţi, folosindu-se FPI-R (Freiburger Personlichkeitsinventares-Fahrenberg, Hempel şi Selg, 1994) şi cuprinde două părţi: 1) o comparaţie între versiunile creion-hârtie şi o versiune prezentată pe computer şi 2) o comparaţie a succesiunii originale şi prezentarea itemilor fiecărei subscale ca bloc omogen, ceea ce presupune extragerea scalelor din ansamblul cuprinzător al inventarului. Rezultatele indică efecte slab pronunţate ale versiunilor cu suport computerizat şi ale succesiunii itemilor (cu referire la valorile medii). Nu se indică nici o interacţiune cu factorul „sex". Parametrii de fidelitate se ridică, în parte, la versiunile computerizate. Prezentarea computerizată duce la o sporire a vigilenţei în timpul efectuării probei. Autoarea consideră că a demonstrat egalitatea dintre FPI-R, standardizată sub forma de aplicare creion-hârtie, şi versiunea cu suport de aplicare computerizat.

Probleme de echivalenţă pune şi un studiu semnat de Jean Roch (1993, pp. 215-219) de la AFPA (Asociaţia pentru Formarea Profesională a Adulţilor). Aceste probleme apar atunci când testele informatizate sunt construite ca forme paralele la testele creion-hârtie. Pentru a putea vorbi despre echivalenţă, trebuie îndeplinite trei condiţii:

- în ceea ce priveşte scorurile obţinute cu cele două forme ale testului, nu trebuie să existe diferenţe semnificative de rang, adică trebuie să se asigure că fidelitatea test-retest este asemănătoare şi cât se poate de ridicată între forme diferite;
- mediile şi varianţele scorurilor nu trebuie să difere semnificativ la cele două sisteme de testare;
- trebuie să se ia asigurarea că nu există interacţiuni între variabile precum anxietatea sau familiaritatea cu informatica, anumiţi indivizi fiind dezavantajaţi, în timp ce alţii sunt favorizaţi când li se administrează un test informatizat.

Dacă aceste condiţii nu sunt îndeplinite, nu înseamnă deloc că trebuie să abandonăm testul informatizat; dimpotrivă, este indispensabil să se modifice normele stabilite pentru versiunea informatizată sau să se ţină cont de anumite variabile moderatoare. Planul experimental a cuprins o baterie de patru teste de capacităţi cognitive şi a fost administrat în 1992 (în cadrul Serviciului de psihologia muncii de la Reims) candidaţilor cu o formaţie profesională de nivel V în sectorul construcţiilor şi al metalurgiei. Bateria cuprinde un test de comprehensiune mecanică (LC), unul de vizualizare spaţială tridimensională (LV), unul de lectură a desenelor foarte simple (LD) şi unul de aritmetică (AR), toate incluse în acelaşi caiet. Variabilele tratate în acest studiu sunt numărul de răspunsuri exacte pentru fiecare dintre aceste teste, precum şi suma celor patru scoruri brute. În versiunea informatizată, itemii sunt prezentaţi individual, subiecţii putând reveni asupra unor itemi şi verifica răspunsurile lor ca în versiunea creion-hârtie. În cele două versiuni, bateria este administrată fără nici o restricţie de timp. Bateria a fost administrată de două ori candidaţilor şi aceştia au fost repartizaţi aleatoriu la una dintre cele patru grupe constituite după ordinea indicată:

Suport test I	Suport test II	Grup	N
Hârtie	Hârtie	PP	49
Calculator	Calculator	OO	49
Hârtie	Calculator	PO	49
Calculator	Hârtie	OP	45

Înainte de începerea testelor de capacităţi cognitive, toţi candidaţii trebuiau teoretic să treacă:

1) un chestionar de 20 de itemi raportându-se la atitudinea în privinţa informaticii;
2) un test informatizat cuprinzând:
   - patru întrebări asupra gradului lor de familiaritate cu maşina de scris, calculatorul, Minitelul şi jocurile video;
   - o frază de 12 cuvinte pe care ei trebuiau să o tasteze repede şi corect;

Rezultatele, prezentate în tabelul următor, indică existenţa unor coeficienţi de corelaţie ridicaţi în sistemul test-retest, mai ales din faptul că retestul urmează imediat testului. Omogenitatea coeficientului de corelaţie test-retest nu pare asigurată (testul 2 de omogenitate a corelaţiilor are semnificaţie la nivelul $p=0,001$ sau $p=0,01$ pentru testele LC, LD şi pentru scorul total). Aceasta se datorează faptului că toate testele şi totalul au o fidelitate mai ridicată când sunt administrate de calculator. Diferenţa se poate atribui calităţilor intrinsece ale suportului informatic sau, mai simplu, împrejurării că pe calculator orice eroare în înregistrarea răspunsurilor (decalaj), ca şi în corectare şi calcul va fi cu siguranţă evitată.

Tabelul 40

Comparaţia rezultatelor la testele cu suport computerizat şi la cele cu suport hârtie

	PP (n=49)	PO (n=49)	OO (n=49)	OP (n=45)	χ^2
LC	0,813	0,850	0,968	0,771	30,08 ($p<0,001$)
LV	0,669	0,753	0,829	0,765	3,27 (N.S.)
LD	0,704	0,897	0,829	0,765	15,48 ($p<0,01$
AR	0,865	0,803	0,866	0,728	4,64 (N.S.)
Total	0,839	0,922	0,971	0,882	20,85 ($p<0,001$)

Legendă: N.S. = nesemnificativ. Statistica $\chi 2$ are asociate 3 grade de libertate.

În privinţa interacţiunii variabilelor, se fac următoarele precizări:
- pentru fiecare variabilă retest şi pentru fiecare individ aparţinând grupelor PO şi OP s-a calculat scorul cel mai probabil după ecuaţia de regresie multiplă, luând cele patru variabile-test ca predictori în fiecare dintre cele două grupe separat;
- în cele două grupe, s-a calculat abaterea între scorurile astfel prezise şi cele realmente observate;
- s-au pus în corelaţie cele patru variabile de deviaţie cu cele cinci variabile de atitudine şi de familiarizare cu informatica.

Abaterile corespunzând erorilor de măsură nu au fost corelate între ele, nici cu alte variabile, cel puţin când exista o interacţiune.

Concluziile au fost următoarele:

1. Testele informatizate de capacităţi cognitive sunt echivalente versiunii lor creion-hârtie de origine şi normele existente nu trebuie modificate. Această concluzie nu trebuie să fie generalizată la proceduri de administrare diferite.
2. În privinţa interacţiunii variabilelor, precum atitudinea în privinţa informaticii, familiarizarea cu calculatorul (şi, indirect, anxietatea), rezultatele nu sunt concludente. Autorul înclină în favoarea unei interacţiuni moderate. Această ipoteză merită să fie verificată într-un studiu ulterior, realizat pe eşantioane mari. Dacă ea se confirmă, rămâne de verificat dacă prognosticul de succes în formarea profesională poate fi semnificativ ameliorat prin luarea în considerare a moderatorilor de acest tip.

Preocupările sistematice pentru introducerea computerizării în psihologie prin intermediul testului psihologic sunt evidente şi în lucrările prezentate la Congresul Internaţional cu tema *Tehnici psihologice de evaluare a persoanelor*, organizat la Paris între 25 şi 27 mai 1993 de INETOP şi EAP. Cu acest prilej, J.A. Fondarai, M. Micas, I. Alma şi P. Avril (1993, pp. 207-214) au studiat utilizarea unor baterii de teste psihometrice informatizate pentru detectarea comportamentului normal şi patologic al indivizilor în domeniul receptării informaţiei. Autorii susţin că evoluţia testului computerizat a luat amploare după anul 1980 şi a permis un larg evantai de cercetări în domeniul psihologiei aplicate. Asimilarea informaţiei se realizează, după părerea autorilor, în trei etape:
- sesizarea propriu-zisă, independent de orice acţiune ulterioară;
- o reacţie nonspecifică, cvasireflexă;
- o reacţie deliberată şi specifică.

În cercetare s-au utilizat trei teste, care au căutat să răspundă la două întrebări:

1. Procentul de reuşită al fiecărui stadiu, al unuia sau altuia dintre acestea, şi timpul utilizat sunt în funcţie de starea mintală a pacientului, dependentă de numeroşi factori (vârstă, patologie, antrenament)?
2. Există o independenţă între rezultatele obţinute prin cele trei teste sau, din contra, un anumit grad de dependenţă? În acest ultim caz, rezultatele diferenţiate permit detectarea punctului de impact al unei patologii în funcţionarea cerebrală?

Pentru cercetare au fost folosite computere MacIntosh sau compatibile MS-DOS. Cele trei teste utilizate se refereau la mascajul vizual, timpul de reacţie şi evidenţierea modificărilor unui pătrat.

• *Mascajul vizual.* Testul a fost conceput şi utilizat de Truvey şi transpus de Debouzy (pe computere Commodore şi Apple) şi de autorii studiului (pe MacIntosh şi compatibile MS-DOS). Elementul esenţial al testului constă în afişarea pe ecran a unei litere de dimensiuni mari pe o durată de 20 ms. Literele au o dispunere spaţială simetrică (A, H, M, O, U, X). Apoi litera se şterge şi urmează o perioadă variabilă de „negru", oscilând între 20 şi 500 de milisecunde, continuată cu prezentarea unei măşti structurate (30 milisecunde). Parametrul luat în considerare este durata „negrului", separând litera de mascaj. După ficcare prezentare, subiectul trebuie să indice ce literă a fost afişată. Se începe printr-o anumită durată a „negrului", dar aceasta este redusă la jumătate dacă pacientul

B.C.U. „M. EMINESCU" IAŞI

răspunde corect de trei ori, fiind refăcută dacă el comite două erori consecutive. Din acest moment, procentajul răspunsurilor bune se înscrie pe o curbă cu alură sigmoidală, al cărei tratament matematic permite liniarizarea şi determinarea unui timp T-50, analog lui DL-50 din farmacologie.

• *Timpul de reacţie* este de o mare simplitate. Sub formă auditivă, aparatul emite la intervale aleatorii un bip. Subiectul răspunde apăsând pe o clapă oarecare. Proba este repetată de un număr de ori reglabil (în general de la 10 la 20). Sub formă vizuală, aparatul afişează o imagine fără semnificaţie egală (cerc sau XXX) la intervale de timp aleatorii şi subiectul trebuie să apese pentru răspuns o clapă oarecare. Se măsoară media timpului de reacţie şi abaterea standard.

• *Testul pătratelor*. În privinţa temei testului, se face precizarea că este inspirat din testul de baraj al lui Zazzo. Se prezintă sub forma a două pătrate, prevăzute cu o coadă dispusă aleatoriu şi care rămân în permanenţă afişate într-un colţ al ecranului, şi a unui pătrat căruia i se trasează o coadă, a cărei poziţie este schimbată la fiecare încercare (100 de încercări). Subiectul trebuie să apese pe litera „O" (pentru „da") atunci când coada figurii centrale este orientată în acelaşi sens ca aceea a unuia dintre pătratele fixe şi pe „N" (pentru „nu") dacă nu se întâmplă acest lucru. Calculatorul evaluează calitatea răspunsului şi timpul scurs pentru obţinerea lui.

Trebuie să semnalăm că solicitările specifice ale testelor de mascaj vizual şi cele ale pătratului pot fi cu greu realizate fără ajutorul computerului. În privinţa subiecţilor examinaţi, aceştia provin din sfera medicinei geriatrice: 280 de subiecţi împărţiţi în trei categorii:

a) subiecţi normali, dar în vârstă (vârsta medie 66 ani);
b) subiecţi prezentând tulburări benigne de memorie (vârsta medie 72,4 ani);
c) subiecţi atinşi de demenţă incipientă (vârsta medie 74,3 ani).

Un alt lot de subiecţi provine din cadrul unei experienţe efectuate asupra parametrilor mintali care pot să condiţioneze calitatea conducerii auto în funcţie de vârstă (650 de subiecţi, dintre care se extrage un subansamblu de 190 din cadrul unei consultaţii, aceştia din urmă fiind normali din punct de vedere mintal şi având o medie de vârstă de 64,6 ani, deci mai puţin decât cei din eşantioanele precedente).

În privinţa concluziilor, se poate spune că performanţele subiecţilor la cele trei teste se disting, pe de o parte, prin factorii de grup (vârstă, patologie şi, desigur, alţii neexploraţi) şi, pe de altă parte, prin susceptibilităţi individuale, oricare ar fi cauzele influenţând una sau alta dintre funcţiile mintale implicate în aceste prime stadii ale asimilării informaţiei. Din acest motiv, testele au putut fi utilizate ca instrumente de detectare a potenţialităţii subiecţilor cuprinşi în domeniul geriatric şi au constituit un instrument diagnostic interesant. De asemenea, utilizarea lor în domeniul conducerii automobilului, în care sesizarea rapidă a informaţiei este în mod firesc importantă, arată că rezultatele variază într-un fel neaşteptat în raport cu vârsta, punându-se deci problema adecvării conducerii auto în funcţie de vârstă. Autorii consideră că testele pot fi relevante şi în alte domenii de evaluare a aptitudinilor mintale.

Jean-Luc Roulin considera, în 1990, că efectuarea automatizată a testelor este puţin reprezentată, cazurile cele mai frecvente fiind chestionarele de interes şi testele de personalitate. El considera, la momentul respectiv, că informatizarea testelor pare să

cuprindă doar probele creion-hârtie sau doar pe acelea în care subiectul are de ales un ansamblu limitat de variante. În comparaţie cu probele clasice, avem un nou mod de prezentare a stimulilor, un alt mod de a răspunde la solicitările acestora şi unul de prelucrare şi raportare a răspunsurilor. O serie de cercetători au încercat să evalueze deformarea răspunsurilor din cauza experienţei anterioare a subiecţilor cu calculatorul. Majoritatea rezultatelor cercetărilor arată că versiunile creion-hârtie şi cele automatizate ale aceluiaşi test nu produc rezultate sistematic diferite. R.J. Sternberg apreciază că există slabe corelaţii între performanţe şi experienţa cu computerul. Utilizarea calculatorului nu pare să aibă aceeaşi influenţă asupra diferitelor categorii de populaţie (apreciate după vârstă, sex, nivel de cultură). Pentru persoanele în vârstă, forma informatizată a unui test pare mai acceptabilă decât forma creion-hârtie.

În privinţa problemelor etice, ne putem orienta după liniile directoare ale APA din anul 1986, care consideră că aceste realizări ştiinţifice pot fi vândute de către institutele private dacă îndeplinesc standardele de competenţă. Problemele de drept ale folosirii sistemului trebuie să se rezume la cerinţa ca specialiştii din institute să fie calificaţi, adică pentru conceperea lor să se apeleze numai la psihologi, psihiatri şi nepsihologi (pentru partea informatică) atestaţi prin diplome universitare.

Opiniile despre validitatea sistemelor (mai ales la producerea expertizelor automatizate) sunt, de asemenea, puternic diversificate. O parte dintre cei ce emit păreri avizate consideră că validitatea este greu de realizat, dată fiind lipsa unei metode adecvate şi, în general, a rezultatelor unor cercetări concordante. Acumulările de creaţii computerizate ale testelor au dus la apariţia unui set de realizări care figurează în mod distinct în lucrările de specialitate consacrate testului psihologic (vezi Gregory, 1996, pp. 574-577).

CBTI (*Computer-based test interpretation*) sunt produse software specializate care se referă la interpretarea testelor sub forma scrierii unor rapoarte de către computer. Serviciile respective sunt accesibile prin e-mail, in/on-line computer, on-site microcomputer. Piaţa pentru astfel de produse este prosperă şi se estimează creşteri accentuate în viitor. În 1987, J.N. Butcher a prezentat o listă de 187 de produse distincte. În 1990, numărul lor crescuse la 408. *APA Monitor* prezintă noutăţile de la lună la lună. Prezentarea lor detaliată ar ocupa paginile mai multor volume. În aceste rapoarte se disting părţi specializate în stabilirea scorurilor, în statistică, în aspecte descriptive şi în domeniul clinic.

În 1984, R.I. Lanyon critică mai ales strategiile de piaţă ale ofertanţilor comerciali şi informaţiile greşite despre validitatea sistemului. Reclama făcută produselor comerciale are scopul clar de a accelera vânzările, dar în domeniul psihologic efectele acestei exagerări a calităţii afectează serios relevanţa informaţională a rezultatelor obţinute prin intermediul lor.

În 1985 şi 1987, K.I. Moreland oferă o imagine generală asupra punctelor critice ale sistemelor MMPI automatizate, insistând asupra carenţelor din cercetările de validitate, şi dă indicaţii practice pentru potenţialii utilizatori ai unor astfel de sisteme. Se subliniază ideea potrivit căreia sistemele au la bază metoda unei interpretări şi prognoze automatizate. Aceasta înseamnă că este vorba numai despre domenii parţiale ale sistemelor de înregistrare. Interpretările se fondează predominant pe experienţa unui expert în diagnostic individual. Studiile de validare empirică folosesc frecvent expertiza tradiţională a experţilor în diagnostic.

4. Sisteme de teste computerizate

O perspectivă cuprinzătoare asupra introducerii diferitelor teste computerizate sau sisteme de teste în diverse domenii de aplicaţie, în SUA, este oferită de J.N. Butcher (1987), care a elaborat un ghid apreciat privind probele psihologice computerizate, ca şi B. Plake şi J.C. Witt. Un raport al produselor software din SUA pentru diagnosticul asistat de computer, cu descrieri scurte, îl oferă catalogul lui M.L. Stoloff şi J.V. Couch (1987), unde sunt recenzate 365 de programe. Acestea cuprind realităţi diferenţiate ale testelor automatizate, intervenţia calculatorului fiind parţială în unele situaţii. Referinţele principale sunt corectarea, interpretarea şi administrarea propriu-zisă a testelor.

În spaţiul german, un diagnostic computerizat se produce, predominant, în cadrul dezvoltării şi introducerii testelor individuale (practic, un test aparativ computerizat) sau a sistemelor de teste computerizate (frecvent, versiuni computerizate ale procedeelor creion-hârtie sau părţi din acestea ; de exemplu, dischete pentru evaluarea rezultatelor la testul Rorschach). Programele sunt adaptate pentru PC, examinându-se reacţiile examinatului, valorizându-se răspunsurile şi producându-se valori normative. În câteva cazuri, rezultatele individuale sau configuraţiile de profil sunt analizate critic şi, eventual, interpretate, fiind reunite într-un raport. La sfârşitul anului 1992, sunt oferite pe piaţa germană cinci sisteme diferite, pentru utilizarea pe un computer personal.

Firma ZAK produce sistemul „Test 3000". Computerul personal serveşte ca instalaţie de conducere a testului, care poate să deschidă până la opt locuri de lucru pentru subiecţi, constând din ecran şi periferice suplimentare, ca, de exemplu, aparate de determinare a reacţiilor.

La fel lucrează cuprinzătoarea versiune PC *Wiener Testsystems* a firmei austriece Schufried. Din locul de conducere a testării pot fi deschise multiple periferice (ecranul candidatului, tastatura candidatului, aparat de reacţie, spot luminos). Oferta de software cuprinde numeroase versiuni ale testelor standard tradiţionale ale editurilor Beltz, Hogrefe, Huber (teste de inteligenţă, teste de personalitate, teste clinice, proceduri pentru diagnosticarea anxietăţii, dispoziţiilor, aptitudinilor, procedee de măsurare fiziologică etc.), programe ale *Curatorium*-ului pentru siguranţa circulaţiei (printre altele, o baterie de teste psihologice pentru circulaţie), numeroase programe ajutătoare şi suplimentare. Firma Hogrefe (centru de aparatură) a dezvoltat un sistem psihodiagnostic sprijinit pe calcul, numit „RPS". La un post de lucru cu tastatură proprie, subiectul poate să realizeze în interiorul unei suprafeţe de utilizare unitară o serie de procedee-test computerizate, care ulterior vor beneficia de un raport.

În fosta RDG, a fost dezvoltat *Leipziger Testsystem* care avea un PC ca loc de conducere a examenului, cuplat, pentru folosinţa examinatului, la calculatoare mici, specifice, şi aparatură periferică. Şi acest sistem de examinare avea la dispoziţie o largă ofertă software.

La Universitatea Humboldt din Berlin a fost dezvoltat sistemul diagnostic *Leila* (K.D. Hansgen, 1991), care între timp a fost comercializat de către editura Hogrefe. Sistemul are destinaţia de a fi introdus în practica de diagnostic curentă şi prezintă procedee de a aborda domenii precum inteligenţa, aptitudinile specifice, randamentul, memoria, atenţia şi concentrarea senzorio-motrice, personalitatea, concepţia despre sine, precum şi chestionare pentru înregistrarea durerilor, simptomelor, intereselor şi tendinţelor.

Cu toate că majoritatea procedeelor integrate prezintă adaptări ale vechilor probe de forma creion-hârtie, între timp au fost produse noi sisteme de normare (de fiecare dată stanine) pentru versiunile computerizate.

Un sistem special a fost produs în Belgia, la Center for Basic Interactive Research in Assessment & Testing. Este vorba despre un sistem în reţea.

Sisteme diagnostice asistate de computer în marile organizaţii. În cadrul marilor instituţii, şi-au făcut apariţia baterii de teste autonome, produse în cadrul unor cercetări reprezentative de mare anvergură, permanent actualizate ; rezultatele au fost utilizate, pe baza unor valori fundamentate empiric, în precizarea diagnostică, consiliere şi clasificare, de către evaluatori cu experienţă.

Un exemplu este furnizat, în Germania, de aplicarea diagnosticului computerizat la Oficiul Federal pentru Muncă (J.F. Booth, 1986 ; H. Eckhard, 1977 ; W. Engelbrecht, 1975, 1977, 1979, 1980). În acest caz, au fost dezvoltate trei sisteme diferite de teste automatizate : *Seria de teste pentru şcoala generală* (EUB), *Seria de teste pentru bacalaureaţi şi studenţi* (MOT) şi *Seria de teste pentru adulţi* (ETS). Sistemul cel mai frecvent utilizat este EUB, cu care au fost examinaţi, începând din 1980, aproximativ 60.000 de elevi. Bateria de teste este utilizată ca procedeu hârtie-creion în oficiile de muncă regionale (durata de realizare : circa o oră). Chestionarele de răspuns computerizabile sunt trimise la sediul central din Nürnberg şi evaluate automat. Se vizează consilierea pentru cei care aşteaptă acest lucru şi, suplimentar, pentru serviciul psihologic al forţelor de muncă. Bateria constă din 8 teste de randament (domeniul aptitudini I), 12 scale pentru autoevaluarea aptitudinilor profesionale (aptitudini II) şi 12 scale pentru aprecierea intereselor profesionale. Valorile-normă originale au fost produse pe baza unui eşantion de 30.000 de şcolari. Pentru validarea bateriei de teste şi pentru producerea profilurilor de referinţă pentru diverse grupe profesionale, au fost realizate două studii *follow-up* (1974 şi 1977), chestionarele fiind apoi trimise participanţilor precedenţi EUB. Pentru fiecare beneficiar care aştepta o consiliere şi participa la cercetarea EUB, era disponibil (cu ajutorul programului) un caiet de evaluare. Caietul cuprinde profilul caracteristicilor celui care aşteaptă sfaturi (aptitudini I, aptitudini II, interese), propuneri profesionale şi eventual, pe baza profilului de interese al beneficiarului, indicaţii cu privire la atitudinile pentru diverse domenii profesionale. Pe baza rezultatelor caietului de evaluare, are loc consilierea subiectului de către un specialist. În cazul în care se constată rezultate cu putere slabă de prognoză (de exemplu, nici o aptitudine reliefată sau nici un interes, nici o propunere profesională sau prea multe) sau în cazul existenţei unor rezultate contradictorii (de exemplu, discrepanţa dintre aptitudinile cuprinse obiectiv şi autoestimare sau interese), solicitantului i se poate recomanda o expertiză psihologică individuală de către serviciul psihologic.

După ce bateria de teste EUB a fost prelucrată şi validată din nou, aceasta a fost desemnată ca BWT (*Berufswahltest*, test de alegere profesională). A fost lansat un nou studiu *follow-up*. Concomitent (începutul anilor '90), s-a întocmit proiectul sistemului *DELTA*, care va permite realizări directe şi evaluări ale diferitelor procedee de testare. Pe de o parte, există un computer central şi, pe de altă parte, mai multe posturi de lucru pentru subiecţi. Este vorba despre un sistem în reţea, unde puteau să funcţioneze concomitent până la 16 posturi de lucru pentru candidaţi, în aşa fel încât diverşi candidaţi să poată efectua diverse serii de teste. Realizările pot fi controlate de la locul special de conducere a testării. Rezultatele testelor pot fi depozitate atât regional, cât şi la centrala

din Nürnberg, şi vor sta la dispoziţie pentru mai multe lucrări de cercetare. La terminalul de referenţi pot să fie produse formulare adiţionale, un pas spre examinare fără o administrare a fazelor tradiţionale (Hagebock, 1994, pp. 42-43).

Bundeswehr este o altă organizaţie vastă, care realiza într-o vreme circa 350.000 de cercetări de aptitudini anual doar în vestul Republicii Federale, în care au fost dezvoltate şi introduse teste computerizate (M.W. Habon, 1991 ; H.U. Nauels şi W. Wildgrube).

Începutul testării aptitudinilor de admitere asistate de calculator se produce în 1983, în primul rând la aşa-numiţii „candidaţi voluntari" din Hanovra şi München, folosindu-se aparate speciale ale firmei ZAK şi Schufried (CAT I). În 1985, are loc o schimbare la computerul personal al firmei ZAK şi se dezvoltă o nouă versiune pentru software de evaluare (CAT II). Noul sistem a fost introdus din toamna lui 1989 la toţi candidaţii voluntari ai Bundeswehr. În versiunea CAT II, au fost realizate în total 11 subteste ale bateriei de testare a aptitudinilor (în versiunea anterioară erau doar 7). Bateria de teste computerizate, cu care au fost examinate până în 1988 cam 25.000 de persoane, a decis mai bine în privinţa candidaţilor decât s-a făcut în forma convenţională (W. Wildgrube, 1990). Aceste rezultate corelează puternic cu felul alegerii specifice a subiecţilor. Este vorba exclusiv despre tineri adulţi bărbaţi, care s-au angajat voluntar în serviciul Bundeswehr. Punctele critice pentru dezvoltarea şi verificarea procedeelor de testare computerizate (printre care realizarea omogenităţii perechilor de itemi, alegerea unui model de estimare adecvat pentru parametrii itemului, stabilirea criteriilor de departajare, estimarea stabilităţii pentru parametrii aptitudinii) au condus de timpuriu la detectarea părţilor dificile în testarea convenţională cu sprijin computerizat. Au fost dezvoltaţi şi evaluaţi, în cadrul unui contract de cercetare cu Universitatea din Aachen, perechile de itemi necesari introducerii algoritmilor adaptativi.

Într-o nouă versiune, CAT III (introdusă din 1990), pe lângă transformări hardware şi software (dezvoltate de firma Dornier), s-a adăugat şi un algoritm adaptativ, ce-i drept, la început numai pentru studii de probă. Diferitele strategii ale testării adaptative trebuiau schimbate şi evaluate, în special în privinţa punctelor critice amintite.

B. Bukasa, S. Kisser şi U. Weninger aduc informaţii despre un sistem de diagnostic computerizat, adaptat cercetării aptitudinilor de psihologia circulaţiei, care a fost dezvoltat în Austria şi care, din 1983, funcţionează la *Curatorium*-ul pentru siguranţa circulaţiei şi pentru expertiza aptitudinilor. Este vorba, în parte, despre noi dezvoltări şi, în parte, despre transpuneri ale testelor tradiţionale. Pentru întregul sistem, au fost elaborate tabele de norme şi studii de validare. Bateria de teste de psihologia circulaţiei şi procedurile suplimentare sunt integrabile în *Wiener Testsystem* şi comercializate prin firma Schufried. Din 1983, sistemul a fost utilizat, în Austria, în mai mult de 50.000 de examinări, fiind între timp preluat de către Centrul de Cercetări Psihologico-Medicale al „Uniunii de supraveghere tehnică Bayern", în vederea aplicării la expertiza circulaţiei.

O experienţă foarte elocventă pentru aprecierea procesului de computerizare în activităţile de evaluare psihologică este cea oferită în Franţa, de SNCF. Încă din 1931, J.M. Lahy creează un laborator psihologic în cadrul Căilor Ferate din Nord. Din 1957, interviul clinic, realizat de un psiholog specializat în psihologia muncii, completează şi îmbogăţeşte testele. SNCF practică evaluarea pentru toate categoriile de personal, pentru toate meseriile şi pentru întregul ansamblu de etape de gestionare a resurselor umane. În coordonarea psihologilor diplomaţi, se realizează anual în jur de 15.000 de bilanţuri psihologice (vezi Macaire, 1994, p. 804). Evaluarea se desfăşoară în patru etape de

bază : metoda testelor, observarea comportamentului, interviul psihologic şi sinteza evaluării. Informatizarea testelor psihologice, la SNCF, a început după anul 1980 şi a înregistrat progrese pe două planuri. Primul plan este cel al calităţii metodelor de evaluare, care cuprinde următoarele aspecte :

- fiabilitatea corectării testelor ;
- punerea la dispoziţia imediată şi prezentarea mai adecvată a rezultatelor examenului ;
- sporirea numărului de indici şi de parametri exploatabili informaţional ;
- îmbunătăţirea condiţiilor de administrare a testelor (prin controlul standardizării, supleţea utilizării şi respectarea ritmului individual, disponibilitatea psihologului pentru observarea comportamentului) ;
- acceptarea socială superioară a testului de către toate generaţiile ;
- reactualizarea mai frecventă a etalonării.

Un alt plan este cel al reducerii costurilor pentru SNCF, ceea ce se concretizează în reducerea timpului de corectare a testelor, micşorarea valorii investiţiilor şi a întreţinerii materialului de testare, reducerea costurilor de realizare a studiilor (privitoare la elaborarea testelor noi, la etalonare, la rezultatele examenelor).

Trebuie remarcat faptul că informatizarea a permis crearea unor teste mai performante, mai ales pentru personalul implicat în siguranţa circulaţiei.

Jean Pierre Macaire, şeful Serviciului psihologic de la SNCF Paris, apreciază că validitatea şi fiabilitatea rezultatelor sunt mult superioare acelora ale testelor tradiţionale. Un alt avantaj se referă la cantitatea de informaţie care este posibil de tratat. Se apreciază că informatica va bulversa domeniul psihologiei muncii, prin crearea unei adevărate bănci de comportamente pentru uzul psihologilor. Toate testele au fost concepute de către sectorul de cercetare al Serviciului psihologic şi sunt integrate într-un sistem de teste informatizate, care, la rândul său, este integrat unui sistem de gestiune computerizată.

5. Testele adaptative computerizate (CAT)

Un loc special în analiza creaţiilor psihologice computerizate trebuie rezervat aşa-numitului *computerized adaptative testing*. Acest sistem presupune stocarea în memoria calculatorului a unui număr mare de itemi diferiţi, a datelor de etalonare, precum şi tratarea informaţiilor în timpul administrării testului (ceea ce se numeşte în termeni de specialitate *sistem de tratare a informaţiei în timp real*). Prototipul unor astfel de teste este scala Binet-Simon, iar modelul de tratare a datelor este cel probabilist, pus la punct de Rasch. Calculatorul determină o ordine de apariţie a unui nou item pe baza evoluţiei subiectului în timpul efectuării probei şi permite o creştere a preciziei măsurării. Există mai multe tipuri de teste adaptative, în care computerul face posibilă dezvoltarea unor algoritmi de mare complexitate pentru selecţia itemilor, pornind de la modele recent elaborate în construcţia testelor destinate selecţiei. Nume de referinţă printre creatorii unor asemenea modele sunt cercetătorii americani J.N. Butcher, L.S. Keller, R.K. Hambleton şi H. Swaminathan.

Testarea adaptativă face apel şi la un nou orizont teoretic pentru procesul de examinare prin intermediul testelor psihologice. Noţiunea de *strategie*, concepută ca un sistem de reguli pentru a ajunge la o decizie, are o importanţă centrală pentru diagnosticul psihologic.

O strategie are menirea de a stabili ce trebuie să facă cel care ia decizia într-o condiţionare posibilă şi, în special, să precizeze ce informaţii îi sunt necesare şi cum trebuie să le valorifice. Strategiile se stabilesc înainte de dobândirea datelor, fiind însă flexibile în funcţie de datele culese şi de subiectul în cauză. Flexibilitatea sau adaptativitatea caracteristice domeniului diagnosticului prin test sunt implementabile în micro- sau macrodomenii (vezi Rettig şi Hornke, 1995, pp. 557-563).

Macrostrategiile presupun stabilirea de reguli pentru ca subtestele, testele şi bateriile de teste să fie selectate şi compuse pentru o anumită problematică diagnostică. Macrostrategiile actuale sunt flexibile, direcţionate spre scopul cercetării şi orientate spre feedback. Frecvent, se aplică o regulă prestabilită în privinţa testelor fără modificări care iau în considerare rezultatele parţiale (intermediare). Aceasta duce, de obicei, la rezultate superficiale ale testului care sunt doar în mică măsură relevante diagnostic pentru luarea deciziilor.

Astfel, strategiile *single-stage* (stabilite de Cronbach şi Gleser, în anul 1965) sunt concepute cu referire la relaţiile etice cu subiecţii şi în raport cu aspectul economic, funcţional al testului. Astfel de strategii, care se afirmă în practică, permit celui însărcinat cu decizia să aibă în vedere condiţiile marginale organizatorice ale examinării, nespecificitatea unor atribuţii şi limitarea deciziei şi acţiunii. Diagnosticul devine mult mai eficient prin utilizarea unor strategii secvenţiale sau *multi-stage*, care permit luarea unor decizii de continuare a testării când ne aflăm în incertitudine în privinţa admiterii sau respingerii unui candidat. Cea mai simplă formă de strategie de testare este strategia *two-stage*. În acest caz, se evidenţiază două reguli de decizie:

- când subiectul în testul A are o valoare de test mai mică decât A_{crit} (valoare critică), atunci acesta este testat cu testul B;
- când subiectul are în testul A o valoare la test mai mare sau egală cu A_{crit}, atunci îl testăm cu testul C.

Regula presupune că valoarea de test la testul A trebuie să existe înainte ca diagnosticianul să ia o decizie strategic flexibilă şi orientată spre investigaţie cu testul B sau C, ceea ce nu este simplu de realizat din punct de vedere strategic fără ajutorul computerului.

Microstrategiile presupun că se iau decizii investigatorii referitoare la o cercetare pe un domeniu de itemi sau, cel puţin, în domeniul unei cantităţi parţiale de itemi care aparţin unui construct acceptabil, respectiv dovedit unidimensional. Luarea deciziilor presupune soluţionarea problemelor de mai jos:

- dacă la un anumit nivel de informaţii, referitoare la o persoană, trebuie să existe un item ulterior sau dacă trebuie să fie întreruptă cercetarea cu o decizie finală;
- cu care item al cercetării trebuie să se continue, atunci când par necesare mai multe informaţii?

Răspunsurile la astfel de întrebări depind de răspunsurile subiecţilor la itemii anteriori, motiv pentru care se foloseşte denumirea de *procedeu de testare dependent de răspuns* (*response contingent*). În cazul unui test convenţional, ambele probleme sunt decise de constructorul testului pentru toţi subiecţii, deoarece sunt prestabiliţi toţi itemii unui test. HAWIK-R este considerat un exemplu de test adaptativ, deoarece anumite subteste sunt adaptate randamentului obţinut anterior. Testele de acest gen îşi găsesc aplicarea în psihofizică, pentru precizarea pragurilor perceptive. Diverse strategii

dependente de răspuns sunt aplicate şi în domeniul programelor computerizate de învăţare (învăţare adaptativă).

Deoarece la testele dependente de răspuns numărul itemilor, alegerea şi succesiunea seriilor de prezentare sunt adaptate la comportamentul de răspuns individual, este greu să se realizeze proceduri de testare de acest gen în cadrul examinărilor de grup şi pe baza testelor creion-hârtie. De regulă, complicatele proceduri de ramificare decizională presupun utilizarea unui microcomputer, ceea ce a dus la crearea domeniului testelor computerizate dependente de răspuns (CAT). La procedele de testare dependente de răspuns, trebuie făcută deosebirea între procedeele secvenţiale şi acele tipuri desemnate în literatura de specialitate, în sens îngust, prin „procedee de testare adaptative". Procedeele secvenţiale sunt cele care trebuie să ducă la decizia tipului de selecţie şi de clasificare şi care se construiesc pe suportul statistic denumit în lumea specialiştilor *sequentiel probability ratio test*. Testele adaptative pot fi introduse pentru acele scopuri care prevăd prioritar măsurarea, pe cât posibil exactă, a caracteristicilor latente şi în relaţie cu deosebirile interindividuale. Testele adaptative pot fi fracţionate în grupe „variabil ramificate" şi „puternic ramificate" şi în teste adaptative ramificate în „subtest--subtest" şi „item-item".

Principiile fundamentale ale testelor adaptative. Testele adaptative ramificate variabil sunt caracterizate prin aceea că ordinea itemilor într-o aplicaţie a testului nu este stabilită *a priori*. În multe cazuri, este căutat individual (pentru fiecare candidat) itemul optimal sub aspectul maximizării informaţiei, dintr-un mare grup de itemi. Procedura este expusă în continuare (*cf.* Rettig şi Hornke, 1995, p. 560).

Testul începe cu un item de start indicat. La fiecare prezentare a unui item este înregistrat răspunsul subiectului şi se face o estimare a aptitudinilor personale pe baza tuturor itemilor prezentaţi anterior până la un anumit punct temporal, precum şi reacţiile observate în această situaţie. În funcţie de un anumit criteriu, se poate aprecia dacă testul trebuie întrerupt sau nu. Dacă testul trebuie să fie continuat, este căutat şi aplicat un nou item în scopul de a îmbunătăţi estimarea aptitudinii. O serie de autori au în vedere posibilitatea de a construi *ad-hoc* următorul item pe baza unor operaţii cognitive cunoscute şi a căror combinaţie se face în raport cu modelul de test logistic liniar al lui Fischer. Reprezentări asemănătoare sunt realizate de diverşi autori în raport cu învăţarea cu ajutorul computerului (exemple concrete se găsesc într-un articol semnat de Stefan Szamosközi, 1997). Ideea testelor adaptative (în terminologia engleză întâlnim frecvent noţiunea de *tailored test*) constă în confruntarea subiectului numai cu acei itemi care corespund nivelului său de aptitudini. Acest mod de testare are avantaje economice, deoarece este evident ineficient să se pună probleme prea simple sau prea grele, ale căror răspunsuri sunt puţin edificatoare pentru nivelul aptitudinal al subiectului. Există şi un avantaj psihologic: prezentarea unor itemi facili poate duce la scăderea concentrării şi a pregătirii pentru randament, iar prezentarea unor itemi prea dificili poate duce la demoralizare şi frustrare şi, în consecinţă, la falsificarea rezultatului.

În privinţa ipotezelor din modelul testelor adaptative, ele sunt realizate în principiu pe baza teoriei clasice, elaborată de Lord şi Novick, în 1968, şi s-au bucurat de un impuls special, prin dezvoltarea modelelor psihometrice cunoscute sub numele de *Item Response Theorie* (IRT). Un model special îşi are originea în scrierile matematicianului danez Georg Rasch. Există însă o mulţime de alternative structurale pentru algoritmul testului, care se deosebesc prin expertiza statistică a estimării aptitudinilor rezultate şi

a rapidității cu care pot fi calculate estimările. Pe lângă aspectele cantitative, trebuie luate în seamă următorii factori de influență: mărimea grupului de itemi, gradul de dificultate al itemilor în grupul de itemi, precizia în estimarea parametrilor itemilor, alegerea itemilor de start, mărimea pasului în faza inițială, parametrii adevărați ai aptitudinii și regulile de întrerupere. Fiecare dintre acești factori constituie obiect de cercetare pentru specialiștii construcției testului. Problemele esențiale rămân alegerea și optimalitatea algoritmului testului.

Punctul de plecare și ipoteza pentru testele adaptative constau în alegerea unui număr suficient de mare de itemi. Spre deosebire de testele convenționale, sunt necesare colecții de itemi constituite din cel puțin 100 de itemi. După stabilirea colecției, o problemă o constituie alegerea unui model IRT, adaptat intențiilor de măsurare și obiectului investigației. Banca de itemi trebuie să fie adecvată modelului ales (verificarea unidimensionalității și a independenței stohastice locale). Dacă este cazul, trebuie selectați câțiva itemi. De asemenea, populația de subiecți vizată trebuie adaptată în așa fel la gruparea itemi. De asemenea, populația de subiecți examinat, suficient de mulți itemi apropiați aptitudinii adevărate. În caz contrar, se poate ajunge la o subestimare sau supraestimare a aptitudinii subiecților. Este de la sine înțeles că, pentru aspirațiile testării adaptative, trebuie să existe și o dotare aparativă adecvată procedurii de testare, în care să fie luate în seamă aspectele fiziologice și psihologice. Asemenea exigențe pot fi satisfăcute deplin doar prin administrarea computerizată a testelor.

Acest sistem de examinare, caracterizat de precizie și eficiență, presupune costuri foarte mari de implementare, motiv pentru care se poate realiza numai în cadrul unor mari instituții. Am amintit într-o secțiune anterioară câteva dintre aceste implementări. Există o serie de detalii diferite de la o metodă de organizare la alta, dar o parte dintre trăsături sunt comune:

a) testele computerizate adaptative (CAT) au o fază de pretestare în care caracteristicile fiecărui item sunt evaluate exact;
b) programul înglobează respectivele caracteristici, precum și o strategie de selecție a itemilor;
c) itemul care urmează este selectat ținând cont de întreaga evoluție a răspunsurilor până la acel punct;
d) nivelul de abilitate estimat al subiectului este recalculat după fiecare răspuns;
e) precizia măsurătorii este estimată după fiecare răspuns;
f) încetarea examinării este condiționată de atingerea unui nivel predeterminat de precizie;
g) punctajul subiectului este stabilit pe baza nivelului de dificultate și a unor măsurători caracteristice ale itemilor efectuați, și nu pe numărul total de itemi corecți.

CAT oferă garanția că fiecare subiect este apreciat cu același grad de precizie, deoarece examinarea continuă până când acesta este atins. Valorile extreme ale testelor tradiționale au niveluri mai mari de eroare decât cele centrale. Robert Gregory atenționează că informația tehnică pentru astfel de probe este greu de obținut. Sistemele CAT reduc lungimea unui test cu aproximativ 50%. Armata SUA și Serviciul pentru educație din același stat (ETS) au în dotare astfel de sisteme de examinare psihologică.

6. Clinic contra actuarial în domeniul psihodiagnosticului

Unul dintre cele mai controversate aspecte ale introducerii computerului în activitatea de psihodiagnostic este cel referitor la elaborarea rapoartelor scrise în urma efectuării unei examinări psihologice sau a introducerii datelor rezultate din corectarea unei probe de forma creion-hârtie. Am amintit anterior tipurile de rapoarte utilizate până în prezent. Cele mai controversate sunt rapoartele descriptive și cele clinice. Aceste rapoarte nu sunt scrise arbitrar de către calculatorul electronic, ci sunt elaborate potrivit implementării programelor software, care se bazează pe recomandările creatorului testului.

Raportul descriptiv merge mai departe decât simpla afișare a scorurilor și redă interpretări sumare ale scalelor unui test (mai ales în cazul celor de personalitate). Utilitatea deosebită a unor astfel de rapoarte este recunoscută mai ales în situația când specialiștii în sănătate mintală au puține cunoștințe amănunțite despre testele utilizate. De exemplu, în cazul chestionarului MMPI se prevede clar ce semnifică un scor înalt la o anumită scală. Prin computerizare nu se falsifică nimic din specificul informațional al testului, deoarece se urmează „conștiincios" precizările din manualul testului, așa cum au fost redactate acestea de autorii Hathaway și McKinley. Pacientul este descris ca anxios, hipersensibil, depresiv ș.a.m.d. Pentru a face exact același lucru fără computer, clinicianului i-ar trebui mult timp pentru a extrage datele de interes pentru pacientul său.

Raportul actuarial depășește simpla descriere și face predicții în ceea ce privește subiectul în cauză. Predicția poate fi implicită (când urmează în mod firesc din acordarea unui diagnostic) sau explicită (prin precizare fără echivoc). Astfel, computerul poate face predicția că un anumit candidat pentru un anumit post (în administrație, de exemplu) se va adapta sau nu exigențelor presupuse ale postului, stabilire care poate fi verificată prin analiza anamnezei (indicând adaptarea la locurile de muncă anterioare) sau prin analiza satisfacerii cerințelor postului vizat. Utilizarea computerului pentru predicție evidențiază distincția esențială între judecata clinică și cea actuarială.

În cazul judecății clinice, specialistul care ia deciziile prelucrează informația potrivit discernământului profesional cu finalitatea diagnosticării, clasificării și prevederii comportamentului.

În aprecierea actuarială, este folosită o formulă derivată empiric pentru a face același lucru, în privința comportamentului. Psihoza poate fi stabilită atât clinic, cât și actuarial (prin introducerea datelor MMPI, de exemplu, într-o formulă bazată pe o cercetare). O formulă actuarială este o ecuație regresivă în care scorurile unui anumit subtest sunt puse în legătură cu un criteriu relevant. În acest caz, reguli statistico-matematice sunt implementate în programele calculatoarelor. Termenii „actuarial" și „computerizat" nu trebuie confundați. Se poate proceda actuarial și fără ajutorul computerului, diferența fiind de economie temporală, și nu de precizie. Pentru a fi actuarială, o interpretare a testului trebuie să fie automatizată, pre-specificată și rutinizată. Robert Gregory (1996, p. 578) se întreabă: „Ce este superior în privința diagnosticului bazat pe MMPI, capul unui psiholog bine pregătit sau o formulă potrivită bazată pe cercetări conștiincioase?". În 1954, Mehl a fost cel care a introdus problematica judecăților clinice contra celor

actuariale în domeniul cercetărilor sociale. El a specificat două condiţii pentru a realiza o comparare justă a celor două modalităţi de luare a deciziilor:

1) ambele metode trebuie să-şi bazeze aprecierile pe aceeaşi informaţie;
2) se impune evitarea condiţiilor care ar afecta în mod artificial acurateţea abordării actuariale. Astfel, ecuaţia actuarială ar trebui să-şi bazeze deducţiile pe un eşantion examinat iniţial.

Pare neverosimil, dar „monstrul actuarial" începe să fie apreciat mai bine decât judecata clinică, în condiţiile în care sunt respectate strict condiţiile funcţionării sale. În 1965, Goldberg a precizat că o sumă liniară simplă de scoruri MMPI dă 70% dintre clasificările corecte, pe când psihologii au obţinut doar 62% în medie (un singur psiholog a obţinut 67%). Robert Gregory, autorul de la care am preluat informaţia, nu oferă informaţii suplimentare în privinţa procedurii de comparaţie. Nu este singura referinţă pozitivă asupra rapoartelor actuariale.

Considerăm că stabilirea valorii rapoartelor actuariale are un drum lung de parcurs, dacă avem în vedere complexitatea şi diversitatea domeniilor psihodiagnosticului contemporan.

7. Preocupări româneşti în domeniul computerizării testelor

Realizările româneşti în materie de computerizare a instrumentelor de investigaţie psihologică au apărut mult mai târziu decât în Statele Unite şi ţările din Europa Occidentală. Acelaşi lucru se poate spune şi despre lucrările ştiinţifice consacrate computerizării în domeniul cercetării psihologice. Vom trece în revistă doar realizările de început în acest domeniu, apărute în deceniile nouă şi zece ale secolului XX.

Alina Mârza şi Liviu Matei (1989, pp. 345-349) realizează o transpunere computerizată a chestionarului Eysenck (EPQ). Limbajul de aplicaţie a fost Basic, cu care, la data respectivă, operau calculatoarele de producţie românească (echivalente Sinclair) TIM-S şi HC-85. Programul conceput pentru EPQ dădea posibilitatea opţiunii pentru administrarea chestionarului şi pentru simpla corectare a acestuia. Prima opţiune presupune începerea examinării prin apariţia pe display-ul computerului a instrucţiunilor de lucru şi a întrebărilor corespunzătoare din chestionar, cu indicaţii privind modul în care trebuie să se răspundă. Programul informatic este conceput în aşa fel încât subiectul trebuie să răspundă la toate întrebările. A doua opţiune aplicativă oferă posibilitatea corectării răspunsurilor pentru un chestionar, deja administrat în forma clasică. În prelucrările realizate după introducerea răspunsurilor în memoria calculatorului, se ţine cont în mod minuţios de instrucţiunile manualului chestionarului, iar notele brute se transformă în cote standard.

În articolul de prezentare a probei computerizate, autorii expun exigenţele care decurg din practica informatizării şi din cea psihodiagnostică (respectarea cerinţelor psihometrice, accesibilitatea pentru un mare număr de subiecţi, prezenţa unor avantaje în comparaţie cu forma obişnuită de examinare, atenţie pentru validitatea de aspect).

Preocupări teoretice şi aplicative în domeniul informatizării probelor psihologice au Monica Albu şi Horia Pitariu. Preocupările teoretice şi particularităţile activităţii

practice de computerizare sunt expuse de cei doi autori într-o lucrare destinată analizei testelor de cunoştinţe (Albu şi Pitariu, 1993).

La o aplicaţie intitulată PSIGEN, autorii utilizează programe executate pe calculatoare cu o putere modestă (CUB-Z), având sistemul de operare CPM (vezi Pitariu şi Albu, 1989, pp. 205-208). Calitatea instalaţiilor informatice existente în acea vreme în ţara noastră îşi pune amprenta asupra realizărilor. Autorii au conceput un program adaptat necesităţilor momentului, adresat psihologilor care doreau să beneficieze de avantajele computerizării, dar care nu aveau cunoştinţe de informatică şi de programare. PSIGEN are două componente: PSI 1 şi PSI 2. PSI 1 poate fi utilizat pentru colectarea de date în vederea prelucrării, având o serie de condiţionări legate de numărul de itemi, numărul de răspunsuri posibile şi punctajele acordate itemilor. PSI 2 a fost prevăzut cu mai multe funcţiuni, permiţând administrarea testelor sau chestionarelor, corectarea, reprezentarea grafică a cotelor standard T pentru anumiţi factori-scală, determinarea claselor de valori pentru unii dintre factorii-scală. Programul cuprinde prelucrări statistice ale rezultatelor testului.

Preocupările de informatizare ale celor doi autori sunt confirmate prin aplicarea şi analiza rezultatelor la Inventarul Psihologic California, CPI, probă de investigaţie de mare complexitate, destinată evaluării comportamentului normal pornind de la manifestările interpersonale şi interacţiunile sociale cotidiene, şi a cărei adaptare la specificul sociocultural românesc a constituit un efort ştiinţific de lungă durată pentru ambii autori (1992, pp. 235-244).

O abordare profundă a problematicii asistenţei acordate de calculator în investigaţia psihodiagnostică este realizată de Ilie Puiu Vasilescu, într-un articol ştiinţific cu referire specială la nivelurile de complexitate (1991, pp. 129-134).

Autorul evidenţiază mai multe tipuri de clasificare a nivelurilor de implicare a programelor computerizate în scopul obţinerii de informaţii relevante psihologic, referitoare la un examinat. Astfel, se disting trei niveluri:

1) nivelul primar este acela în care implicarea calculatorului este apreciată ca superficială, practic despuind răspunsurile subiectului;
2) nivelul secundar de intervenţie a calculatorului este realizat prin crearea unor Probe Interactive Computerizate (PIC). Confruntat cu exigenţele unor astfel de probe, examinatul depinde, în răspunsurile sale, de răspunsurile anterioare, ceea ce implică o solicitare deosebită;
3) al treilea nivel este unul care realizează atât interpretarea, cât şi redactarea automată a concluziilor rezultate din aplicarea probei. Nivelul al treilea are elemente comune cu sistemele-expert.

Cele trei niveluri necesită programe sofware cu grade de complexitate crescânde, în privinţa competenţei metodologico-psihologice şi informaţionale.

Autorul prezentei lucrări, în colaborare cu psihologul Cornel Havârneanu, a proiectat, în cadrul Laboratorului de psihologie rutieră IJTL Iaşi, o serie de teste psihologice computerizate, fapt care a permis elaborarea unor opinii în acest domeniu (vezi Stan şi Havârneanu, 1989). Cercetările (derulate în 1988-1989) au fost făcute cu ajutorul unor microcomputere TIM-S (echivalente Sinclair de 64 K, dintre care 48 RAM). La majoritatea programelor realizate, s-au avut în vedere modelarea unor solicitări senzorio-motorii – care ocupă un loc important în ansamblul exigenţelor psihogramelor din domeniul

transporturilor rutiere – şi elaborarea unor probe de memorie topografică. Programele au fost prevăzute cu stabilirea automatizată a etalonului şi cu ilustrarea grafică a condiţiilor de adaptare la solicitările testului (în faza de învăţare). Pentru a putea relaţiona mai bine datele obţinute la examenul psihologic cu anumite criterii, s-a creat un fişier computerizat al accidentelor de circulaţie din anii 1984-1988. În câmpurile fişierului am înregistrat date despre locul accidentului, cauzele lui, ora producerii, numărul de ore lucrate în ziua respectivă de conducătorul mijlocului de transport până în momentul accidentului. Fişierul de date era în legătură şi cu un fişier de comenzi, astfel încât, în momentul efectuării unui examen, se aflau la dispoziţie şi o situaţie documentată asupra numărului de accidente ale celui examinat, şi o descriere a lor.

O probă cu aplicare practică şi elaborare statistică adecvată a fost elaborată în cadrul Laboratorului de psihologie IJTL Iaşi de prof. univ. Andrei Cosmovici, probă intitulată RCC. Ea era destinată diagnosticului reacţiilor în situaţii conflictuale şi sub presiunea timpului (vezi Cosmovici, Havârneanu şi Stan, 1992).

În urma acestei practici de proiectare şi realizare de teste psihologice computerizate, am evidenţiat o serie de criterii care disting avantajele folosirii calculatorului în examinarea psihologică, în comparaţie cu procedurile tradiţionale. La nivelul anului 1989, considerăm aceste avantaje ca fiind centrate pe următoarele criterii:

1. *Criteriul timp*. Testele psihologice computerizate introduc o condensare temporală, care nu rezultă din scurtarea timpului de solicitare în timpul probei, ci din modul operativ de prelucrare, afişare şi tipărire a rezultatelor. Acest timp câştigat prin procedurile computerizate de examinare poate fi alocat prelungirii convorbirii psihologice cu subiectul aflat în situaţia de examinare. La timpul respectiv, emiteam ideea potrivit căreia calculatorul nu se interpune între examinat şi examinator, ci oferă posibilitatea prelungirii sensibile a contactului uman direct, atât de necesar acordării unui diagnostic competent.
2. *Criteriul mobilităţii constructive*. În comparaţie cu unele teste de reactivitate senzorio-motorie utilizate până la acea vreme, constructorul poate dinamiza modul de apariţie a stimulilor. Posibilitatea utilizării unor stimuli perturbatori are o gamă largă de aplicare. În limita produselor software şi hardware utilizate, se poate realiza o interacţiune complexă subiect – calculator. Modul în care subiectul, deci un utilizator nespecializat, amator, răspunde situaţiei problematice apărute pe display constituie obiect de preocupări ştiinţifice (vezi, de exemplu, Van der Veer, 1990).
3. *Criteriul particularizării şi individualizării examenului*. În formele de examinare computerizate, se pot efectua comparaţii rapide între date, se pot interpreta nuanţat rezultatele.

Capitolul V

CRITICA TESTULUI PSIHOLOGIC

1. Critici şi negativişti

Pe parcursul evoluţiei sale, testul psihologic a dat naştere unei serii de reacţii negative venite din direcţii diferite, reacţii care trebuie privite cu atenţie şi înţelegere de cei ce doresc să aibă o imagine adecvată în raport cu complexitatea implicaţiilor practicii sale şi cu necesitatea perfecţionării sale. Acestea sunt normale, dacă avem în vedere faptul că testul, ca orice realizare omenească, are imperfecţiuni care trebuie remediate (prin feedback-ul firesc al teoreticienilor şi constructorilor), dar nu trebuie să ignorăm că o serie de critici depăşesc graniţa bunei-credinţe, constituind ostilitate oarbă, lipsită de discernământ şi de informare corectă, proiectând în test anxietăţi de grup, prejudecăţi sociale şi rivalităţi ideologice. Vom încerca o abordare lucidă a acestor critici, pertinente şi mai puţin pertinente, combinând în intervenţiile noastre cunoştinţe teoretice, evidenţiind reacţiile practice şi intersectările unor condiţionări exterioare, ambientale, de natură socială, politică, economică etc.

Reacţiile critice faţă de testul psihologic sunt îndreptăţite în măsura în care ele semnalează lipsuri reale în construcţia şi exploatarea acestuia. Deci criticile oneste ale testului psihologic trebuie delimitate de atitudinile nihiliste, care pur şi simplu încearcă să excludă această metodă din rândul metodelor de investigaţie psihologică. Indiferent de direcţia şi natura unor astfel de atitudini, testul psihologic continuă să se dezvolte, deoarece el nu poate fi anihilat administrativ. O parte dintre aceste atacuri nu sunt fundamentate, fiind expresia insuficientei înţelegeri şi cunoaşteri nuanţate a testului psihologic. În parte, cele justificate vin ca o consecinţă a utilizării testelor de către neprofesionişti, lipsiţi de cea mai elementară pregătire în aplicarea şi exploatarea informaţională, ceea ce explică deprecierea temporară a metodei testului. Jean-Luc Roulin apreciază că despre testul psihologic pot să scrie numai persoanele specializate, care l-au practicat timp îndelungat şi au avut posibilitatea de a cunoaşte „pe viu" dificultăţile presupuse de acest instrument de investigaţie psihologică. O documentare solidă reprezintă, de asemenea, o cerinţă stringentă.

Pe parcursul evoluţiei sale, testul psihologic a întâmpinat reproşuri, cum am menţionat deja, venite din mai multe direcţii.

1.1. Critici din partea psihologilor

Testului psihologic i s-a reproşat, în principal, inadecvarea la destinaţia psihodiagnostică, cu referire şi la fundamentele teoretice. Atitudinile de contestare a valorii testului psihologic au apărut imediat după intrarea acestui instrument de investigaţie în practica psihodiagnostică. Primirea „botezului social" a însemnat pătrunderea sa în practicile procedurale ale instituţiilor sociale din domeniul învăţământului, sănătăţii şi organizării muncii. Apariţia primelor instrumente de măsurare a funcţiilor psihice elementare, a aptitudinilor şi a „reginei" aptitudinilor, inteligenţa, a provocat un deosebit entuziasm intelectual şi a trezit speranţa în posibilitatea de a spori rolul ştiinţei în ameliorarea condiţiei umane.

A. Binet şi Th. Simon au sesizat printre primii faptul că testul psihologic are o putere diagnostică limitată şi că forţarea acesteia poate avea consecinţe neplăcute pentru imaginea psihologului şi a psihologiei.

Pasionaţii celor dintâi aplicaţii au suferit o profundă dezamăgire la apariţia neconcordanţelor între speranţe şi aşteptări. Unii dintre elevii lui James McKeen Cattell şi-au abandonat profesiunea în urma obţinerii unor corelaţii foarte slabe la probele utilizate cu anumite criterii şi, în consecinţă, s-au dedicat altor preocupări ştiinţifice.

Neconcordanţele dintre rezultatele testelor şi cele ale criteriilor practice de departajare a subiecţilor au dat naştere neîncrederii în puterea lor de prognoză. Gheorghe Zapan evidenţiază două carenţe în faza de început a psihologiei aplicate, în legătură cu aplicarea testelor. În primul caz, se descriu rezultatele obţinute de H. Hildebrandt în examenul de selecţie a ucenicilor de la firma Borsig din Berlin (în anul 1927), prin folosirea unor teste cunoscute asupra aptitudinilor: „Aprecierea mărimilor, destoinicia mişcării mâinilor, inteligenţa generală, spirit tehnic şi reprezentare spaţială" (vezi Zapan, 1984, p. 105). Când responsabilii cu formarea profesională din firmă au căutat să ordoneze ucenicii conform capacităţilor dovedite în atelier (după o jumătate de an de la examenul psihologic prin intermediul testelor psihologice), s-a observat că această clasificare este surprinzător de puţin asemănătoare cu cea realizată prin teste, mai ales pentru meseria de lăcătuş. Ulterior, valorile prognostice ale testelor pentru aprecierea mărimilor şi destoinicia mişcărilor mâinilor au fost nule sau negative (corelaţii de -0,22 şi -0,18). Cercetările lui H. Rupp, cunoscut teoretician al selecţiei şi orientării profesionale din epoca respectivă, au găsit rezultate contrare în cazul selecţionării ucenicilor la firma Siemens Schuckert din Berlin. Referindu-se la acest prim aspect, Gheorghe Zapan afirmă: „De aici se vede că noţiunile *aprecierea mărimii, destoinicia mişcărilor mâinilor* sunt prea vagi şi nu includ în ele şi nuanţele de deosebiri ale determinanţilor unei aptitudini generale sau profesiuni date" (*ibidem*).

Al doilea caz relatat de Zapan este incidentul petrecut la Moscova, în anul 1922, la Conferinţa asupra circulaţiei tramvaielor, care a trimis zece candidaţi Laboratorului de psihologie industrială de pe lângă Comisariatul Muncii, pentru a-i examina în vederea stabilirii aptitudinilor sau inaptitudinilor pentru meseria de vatman. Membrii conferinţei se înţeleseseră de la început să trimită două grupe polarizate de câte cinci vatmani (cei mai buni şi cei mai slabi). S-a intenţionat să se examineze psihotehnicienii care făceau predicţii şi, în felul acesta, să se creeze o impresie clară asupra încrederii care ar trebui să li se acorde. Doar unul dintre candidaţii cu aptitudini dovedite în activitatea practică

a fost apreciat de către psihotehnicieni. Nivelul aptitudinal al celorlalţi nouă a fost prognozat contrar realităţii: cei buni au fost declaraţi slabi, iar cei slabi - buni.

R.M. Yerkes scoate în evidenţă o serie de neconcordanţe între aşteptări şi realitate, în raportul redactat în urma examinărilor psihologice realizate pentru armata americană, în timpul primului război mondial. Astfel, în 1921, el publică un raport pe 890 de pagini, prezentând rezultatele analizelor efectuate pe recruţii examinaţi în 1917. Multe dintre aceste rezultate privesc compararea grupelor. Diferenţele de inteligenţă dintre grupe sunt explicate de Yerkes însuşi şi, mai mult încă, de comentatorii raportului său, univoc, prin factori ereditari, pe când explicaţiile ambientale sunt adesea mai mult decât convingătoare.

Două exemple sunt edificatoare pentru analizele ulterioare ale altor psihologi. S-a constatat că o eficienţă intelectuală mai redusă la soldaţii atinşi de paraziţii intestinali (ankilostom) nu este explicată prin influenţa bolii asupra aptitudinilor sau printr-un efect al apartenenţei sociale asupra sănătăţii şi asupra aptitudinilor, ci se preferă să se considere că aptitudinile slabe antrenează condiţii de viaţă dificile, care vor determina apariţia bolii; cu alte cuvinte, slăbiciunea intelectuală este cauza stării de sănătate precare, şi nu consecinţa sa. De asemenea, s-a constatat că negrii din statele americane din nord aveau scoruri mai bune decât negrii din statele din sud (analizele ulterioare diminuează semnificaţia acestei diferenţe). În confruntarea cu respectiva stare de fapt, este emisă teza migraţiilor selective (doar cei mai inteligenţi ar fi emigrat în nord), dar e evident că se poate da o explicaţie mai realistă şi mai fundamentată prin raportarea acestui rezultat la condiţiile de mediu şi de şcolarizare ale celor din sud, unde prejudecăţile rasiale erau mai frecvente şi mai intense.

O critică documentată, dar foarte exagerată (prin selectarea doar a acelor fapte şi evenimente care favorizează atitudinea demolatoare), a testului psihologic şi a caracteristicilor acestuia întreprinde Maurice de Montmollin, într-o lucrare consacrată mai ales domeniului psihologiei industriale (1972). Toate demersurile psihologului industrial pentru a-şi practica profesia sunt „desfiinţate" cu o rară vehemenţă (stabilirea profilului postului, profilurile psihologice individuale, validitatea, eficacitatea etc.). El afirmă că disciplina psihologiei industriale nu-şi justifică existenţa. Montmollin consacră o parte din criticile sale modului de efectuare a examenului psihologic, în special fundamentelor lui noţionale. Conceptualizările aptitudinilor, în calitate de caracteristici stabile ale persoanei, sunt doar nişte „comodităţi verbale, lipsite de valoare practică operaţională. Iată de ce, chiar de la început, psihologii au trişat. Ei procedează ca şi cum descrierea posturilor de muncă se poate face spontan, fără probleme, cu aceiaşi termeni (aptitudini, capacităţi, trăsături de personalitate) pe care ei îi utilizează pentru a descrie omul. Această înşelătorie a fost indusă de simţul comun" (*ibidem*, p. 22).

După ce ridiculizează psihogramele, factorii psihologici, schemele explicative simpliste, Montmollin conchide: „Este imposibil să se descrie munca în aceiaşi termeni psihologici care sunt utilizaţi pentru descrierea candidaţilor. Nici o psihologie industrială nu poate să servească la stabilirea prognosticului asupra reuşitei muncii" (*ibidem*, p. 31).

Criticile venite din partea psihologilor sunt o reacţie la devierile în practică ale testelor şi dezbaterilor asupra utilităţii lor.

Utilizarea masivă a testelor colective, fragilitatea fundamentelor lor teoretice, concepţia unidimensională a inteligenţei, credinţa larg răspândită a ineităţii capacităţilor observate nu pot să conducă decât la devieri în practică.

Testele colective, dacă permit efectiv să se descrie populaţii, sunt evident puţin adaptate pentru sprijinul deciziei privind indivizii, mai ales când ele sunt realizate pe o singură

dimensiune. Or, asemenea teste sunt destul de des utilizate pentru a fonda decizii de afectare (alocare a personalului). Contextul performanţei şi elementele susceptibile de a o explica (care pot fi luate în considerare cel puţin parţial, în studiul individual) sunt ignorate. O metodologie psihometrică în mod competent construită permite evaluarea erorilor pe care subiectul riscă să le comită, dar ea a fost rar utilizată, cu toată rigoarea sa, în practica psihologică.

Absenţa fundamentării teoretice duce la privilegierea criteriilor de validare externă. Se consideră astfel că testul măsoară în mod adecvat inteligenţa, nu pentru că el corespunde cu ideile fundamentate pe care constructorul le are asupra inteligenţei, ci pentru că rezultatele sale concordă cu judecăţile sociale asupra indivizilor. S-a ajuns astfel la definirea inteligenţei prin criteriul adaptării sociale şi la asimilarea rapidă a variabilităţii inteligenţei, cu câteva excepţii, în funcţie de clasamentele sociale. Reuşita în practică, într-o profesie, într-o anumită activitate, este asimilată cu posesia inteligenţei de către individ, calitate denumită adesea şi *inteligenţă socială*. Concepţia unidimensională a inteligenţei vine să întărească această tendinţă. Este cu siguranţă mult mai uşor să se asimileze reuşita şcolară şi competenţa intelectuală, când nu există decât un fel de competenţă intelectuală. Faptul că această competenţă se exprimă în termeni de QI, în factor *g* sau chiar în termeni de stadii de dezvoltare este unul secundar. Din acest punct de vedere, Alfred Binet conştientizase importanţa chestiunii, atunci când compara rezultatele testelor pe elevii buni şi pe cei slabi. El recunoaşte că aprecierile profesorilor „erau îndoielnice" şi observă atent „punctul slab al cercetărilor sale", fixându-şi drept obiectiv definirea inteligenţei cu propriile sale mijloace. Câteva decenii mai târziu, David Wechsler nu va avea acest gen de nelinişte. El defineşte inteligenţa ca „o capacitate globală sau complexă a individului de a acţiona într-un scop determinat, de a gândi raţional şi de a avea raporturi utile cu mediul său" (*apud* Huteau şi Lautrey, 1995, p. 40). Wechsler consideră că validarea finală a scalelor de inteligenţă „vine în principal ca urmare a faptului că ele au dat rezultate bune în practică" (*apud ibidem*), aceste „rezultate bune" semnificând pronosticuri bune ale judecăţii profesorilor din şcoală, ale şefilor de personal, ale agenţilor de formare în industrie şi ale ofiţerilor în armată. Wechsler sublinia ambiguitatea acestui tip de validare: „Testele au fost acceptate în armată şi industrie pentru că, atunci când se făcea bilanţul, nu se putea avea întotdeauna încredere în estimările ofiţerilor şi ale şefului de personal, apoi testele au fost vândute pentru că rezultatele obţinute corepund bine judecăţilor practicii" (*apud ibidem*, p. 41).

Rezultatele testelor de inteligenţă nu reproduc decât imperfect clasamentele sociale, mai ales deoarece ele se raportează esenţial la caracteristicile psihologice cognitive. Aceste rezultate aduc, aşadar, o informaţie adiţională. Este întotdeauna interesant să se constate că elevii cu reuşite şcolare mediocre pot avea totuşi rezultate bune la teste şi este întotdeauna util să se înţeleagă raţiunile acestei discordanţe. Practica testelor pe scară largă are adesea drept obiectiv detectarea unor astfel de cazuri. Dar acest obiectiv are cu atât mai multe şanse de a fi valabil şi frecvent atins cu cât eficacitatea intelectuală va fi descrisă independent de eficienţa socială. Relaţiile vor fi foarte probabil remarcate, dar nu au putut fi incluse în definirea inteligenţei. În caz contrar, măsura inteligenţei riscă să se reducă la o operaţie de legitimare a clasamentelor sociale, iar această legitimare va fi cu atât mai îndreptăţită cu cât diferenţele observate vor fi tot mai mult considerate „naturale", adică transmise ereditar.

Devierile în aplicarea testelor apar, în fapt, ca o consecinţă a îmbinării a trei condiţii:
- o definiţie psihologică deficitară a inteligenţei;
- o reprezentare unidimensională a funcţionării intelectuale;
- o explicaţie a diferenţelor individuale prin ereditatea biologică (vezi *ibidem*).

Ele nu caracterizează deci testele în general, ci anumite teste la un moment şi într-un context dat.

Făcând referiri la domeniul teoretic, amintim faptul că, în anii '30, exista deja o critică a testelor, în cadrul psihologiei, în special a fundamentelor ştiinţifice ale testelor psihologice, critică avându-şi sursa în behaviorism şi în psihologia piagetiană. Pentru primul curent, trăsăturile intelectuale, precum inteligenţa, sunt abstractizări mentaliste cărora li se neagă generalitatea, nu numai pentru definirea formelor de inteligenţă cum fac factorialiştii, ci şi pentru focalizarea asupra relaţiilor dintre răspunsuri şi situaţiile specifice. Dată fiind lipsa de interes pentru procesele superioare, psihologii behaviorişti nu au perseverat în aceste critici.

Criticile formulate faţă de teste de către psihologii Şcolii din Geneva şi de către Piaget însuşi sunt de o altă natură. După al doilea război mondial, o dată cu dezvoltarea psihologiei cognitive, critica psihologică a testelor de inteligenţă se manifestă mai puternic decât în perioada precedentă. Testelor li se reproşează simplitatea ambiguă, considerându-se că măsoară doar performanţele şi nu explică aproape nimic cu privire la mecanismele psihologice, care au dus la producerea lor. Ca atare, testele nu pot să furnizeze decât rezultate dificil de interpretat, încărcate de erori cărora li se ignoră natura şi amploarea. Dezbateri asupra originii diferenţelor individuale şi asupra *biais*-urilor sociale în teste au avut loc şi în alte ţări. În Franţa postbelică, în lucrarea *Théorie de l'orientation professionnelle* (1945), Pierre Naville făcea o critică riguroasă a noţiunii de *aptitudine*. Adoptând teza behavioristă, el insistă asupra plasticităţii aptitudinilor şi asupra importanţei fenomenelor de învăţare şi de adaptare. Critica nu vizează în mod direct testele, punând totuşi sub semnul întrebării interpretările cele mai frecvente ale performanţelor observate.

În urma acestor critici, asistăm la dispariţia aproape totală a preocupărilor psihometrice din cursurile de psihologie ale multor universităţi franceze din anii '60. Acest fapt se explică nu numai prin îndoielile privind avantajele sociale ale utilizării testelor, ci şi pentru că testele apar teoretic ca fiind în mică măsură fundamentate şi modernizate, în comparaţie cu alte domenii ale precupărilor din cercetarea psihologică. O situaţie „tipic franceză" a luat naştere prin reducerea ariei de folosire a testelor, prin diminuarea considerabilă a spaţiului alocat, în educaţia şcolară, construcţiei şi exploatării testelor psihologice, situaţie care a durat însă o perioadă scurtă de timp. Aceste critici interne psihologiei sunt perfect justificate, dar mulţi dintre autorii lor nu au prevăzut toate implicaţiile. Este adevărat că psihologia cognitivă, în numele căreia se critică testele, este mai evoluată decât psihologia inteligenţei, care fundamentează testele de inteligenţă. Dar din această limitare temporară a calităţii teoretice s-a tras concluzia că testele, în general şi în esenţa lor, sunt lipsite de utilitate. O apreciere mult mai pertinentă ar fi aceea că sunt necesare eforturi competente de reactualizare a testelor şi de construire a altora noi; mai precis, este necesar ca, în construcţia lor, să se plece de la cunoştinţe de psihologie adecvate, actualizate (ne referim în special la cunoştinţele de psihologie cognitivă, care au pătruns, ulterior criticilor, masiv în lumea constructorilor de teste). Ideea depăşirii ştiinţifice a testelor a fost susţinută de mulţi psihologi, care au judecat

îngust posibilitatea constructorilor de teste de a asimila rapid noutăţile teoretice din domeniul psihologiei. Cea mai bună dovadă o reprezintă faptul că, ulterior acestor critici, au apărut şi testele piagetiene, a căror originalitate constă în aceea că sunt fundamentate printr-o teorie net mai elaborată de dezvoltare a inteligenţei, teoria stadiilor de dezvoltare. A descrie bogăţia de idei a studiilor piagetiene nu figurează însă printre obiectivele lucrării de faţă. Menţionăm doar faptul că aceste teste, atât cele cu administrare colectivă, cât şi cele cu administrare individuală, prezintă dificultăţi în ceea ce priveşte standardizarea.

Fiind fondată pe raţionalitatea psihologiei cognitive, critica psihologică a testelor vizează exclusiv testele de inteligenţă şi nu cuprinde modurile de evaluare a inteligenţei care sunt calificate frecvent ca iraţionale (morfopsihologia, de exemplu). Nefiind obiect de critici, aceste moduri de evaluare sunt relativ valorizate şi găsesc astfel o condiţie favorabilă a dezvoltării lor. Se pot evidenţia şi alţi factori care explică succesul modurilor de evaluare iraţională. Este surprinzătoare constatarea că aceste moduri de evaluare iraţională sunt dezvoltate în acelaşi timp cu criticile la adresa testelor de inteligenţă.

1.2. Critici din partea altor reprezentanţi ai ştiinţelor sociale

Alte critici aparţin, mai ales, sociologilor care se ocupă de teoria educaţiei, evidenţiind în conţinutul faptului psihologic manifestări ale unor influenţe sociale. Printre criticii testului, se numără nume de prestigiu din psihosociologia şi sociologia contemporane. Ne referim la William White, autorul lucrării *Omul organizaţional*, şi Pitirim Sorokin, autorul cărţii *Fads and Foibles of Modern Sociology* (1956). Dar criticile lor nu se referă numai la teste, ci la un întreg ansamblu de stări de lucruri din civilizaţia occidentală.

O critică veche, care persistă până în zilele noastre, le reproşează testelor că au răspunsuri distorsionate de factori de natură socială (*biais* social; reamintim că termenul *bias* se referă la *deformarea rezultatului unui test prin apariţia unor erori sistematice*), care constau din reflexe datorate particularităţilor sociale de existenţă. Ca atare, se susţine că testul evaluează de fapt trăsături culturale. Astfel de opinii circulă şi în rândul psihologilor. După părerea psihologului francez Henri Salvat, răspunsurile la testele de inteligenţă sunt mai ales „sociale şi culturale", putând fi considerate aşadar *teste de proiecţie*.

Grafic, situaţia de *biais* se prezintă astfel:

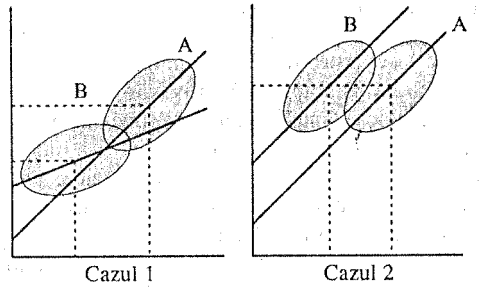


Figura 16. *Reprezentarea spaţială a situaţiilor de* biais

Pe abscise se află rezultatele la test, iar pe ordonată rezultatele la criteriu. În cazul 1, se observă că grupul A diferă de grupul B atât după rezultatele la test, cât şi la criteriu. În acelaşi timp, relaţia dintre criteriu şi test diferă în cele două grupuri (tendinţe de pante diferite). Pentru cazul 2, se observă că există doar o diferenţiere după rezultatele la test (grupul A are rezultate mai bune), în timp ce relaţia dintre criteriu şi test este aceeaşi (tendinţe de pante egale).

S.L. Pressley oferă, în lucrarea *Psychology and the never education*, câteva exemple din practica aplicării testelor (*apud* Pavelcu, 1972, p. 93).

> Exemplul dat de autor se referă la examinarea unui băiat de 12 ani, dintr-o regiune muntoasă, unde şcolile sunt rare; copilului i se aplică versiunea americană a testului Binet-Simon. Prezentăm, în continuare, o suită de întrebări şi răspunsuri la test.
> *Întrebarea nr. 1 adresată copilului:* „Dacă, având 10 cenţi, cumperi bomboane de 6 cenţi, câţi cenţi îţi vor rămâne?".
> *Răspunsul copilului:* „Nu am avut niciodată 10 cenţi şi, dacă i-aş fi avut, nu i-aş fi cheltuit pe bomboane; mama lucrează în fabrică!".
> *(Psihologul continuă cu un alt exemplu.)*
> *Întrebarea nr. 2 adresată copilului:* „Dacă ai dus la păscut 10 vaci ale tatălui şi dintre acestea s-au rătăcit 6, câte vaci ai adus tu în grajd?".
> *Răspuns:* „N-avem vaci şi, dacă le-am avea şi aş pierde şase, n-aş îndrăzni să mă întorc acasă".
> *(Psihologul schimbă din nou exemplul.)*
> *Întrebarea nr. 3 adresată copilului:* „Dacă, din 10 elevi dintr-o clasă, 6 sunt absenţi din cauza rujeolei, câţi elevi vor rămâne în clasă?".
> *Răspuns:* „Nici unul, deoarece celorlalţi le-ar fi frică să nu se îmbolnăvească de rujeolă".

Asemenea tipuri de răspuns apar rar, sunt nişte excepţii, dar ele sunt posibile. În experienţa profesională, am întâlnit un singur răspuns de acest gen, la examinarea cu proba WISC a unui elev de 13 ani de la o şcoală din cartierul ieşean Tătăraşi:

> *Întrebarea psihologului (viza un răspuns din care să rezulte dacă elevul a înţeles raţiunea alegerii deputaţilor într-un stat):* „De ce sunt aleşi deputaţii?".
> *Răspuns:* „Deputaţii sunt aleşi pentru că sunt puşi pe listă, deoarece dacă nu sunt puşi pe listă nu-i alege nimeni, vă garantez eu!".

O altă sursă de deformări ale răspunsurilor o reprezintă proveniența socială a subiecţilor. Primele aplicaţii ale testelor vor pune în evidenţă o inferioritate medie a copiilor proveniţi din păturile sociale defavorizate. Acest fapt a fost explicat plecând de la factori genetici şi ambientali sau plecând doar de la factori ambientali (Reuchlin, 1972). După al doilea război mondial, sub influenţa lui A. Davis, un sociolog de la University of Chicago, a fost propusă o a treia ipoteză explicativă, punând în cauză semnificaţia însăşi a diferenţei observate. Pentru A. David şi colegii săi, K. Eels şi R.J. Havighurst, mai ales testele de inteligenţă avantajează copiii din clasele sociale superioare, căci ele sunt

mai apropiate prin conţinutul lor de cultura acestor clase decât de aceea a mediilor populare. Crezând că se măsoară inteligenţa, se măsoară, de asemenea, distanţa conduitelor care mărturisesc mai mult o apartenenţă socială decât o eficienţă intelectuală. Convingerea acestor sociologi, şi nu numai a lor, este că răspunsurile la solicitările testelor sunt distorsionate social în detrimentul copiilor din mediile populare. Acest punct de vedere nu lipsea nici din criticile formulate la adresa testelor în anii '30, dar ele erau atunci foarte puţin dezvoltate. În concepţia lui Davis şi Eels, valoarea predictivă a testelor avantajează în mod arbitrar copiii din mediile sociale privilegiate (cei doi chiar vor construi un test pe care îl consideră „echitabil din punct de vedere social"). Dar deficienţele de acest gen nu trebuie puse în mod necesar pe seama testelor.

Lucrările lui Davis şi Eels marchează debutul reflecţiei asupra distorsionării rezultatelor din cauza unei influenţe sistematice de natură culturală asupra testului. Psihologii americani au fost nevoiţi să acorde o atenţie foarte mare acestei chestiuni şi, adesea, li se cere câteodată, în faţa organizaţiilor profesionale, să aducă dovezi în favoarea caracterului nearbitrar discriminativ al testelor lor. Testele suspecte de a fi distorsionate din punct de vedere social au fost scoase din uz.

1.3. Critici din partea reprezentanţilor societăţii civile

Diverşi reprezentanţi ai societăţii civile (publicişti, membri ai organizaţiilor neguvernamentale, personalităţi culturale) au acuzat, explicit sau implicit, inegalitatea şanselor în faţa examenului psihologic prin intermediul testului psihologic. În Franţa, numere întregi ale revistei *La Raison* (1952, 1955) conţin atacuri la adresa testelor. În articolele respective, se afirmă că testele psihologice nu răspund nici unei nevoi sociale reale, nici unei instituţii, procedând în folosul particular al lumii capitaliste, constituind nu numai o suprastructură a acesteia, ci şi unul dintre mijloacele sale de presiune, de agresiune şi de propagandă. Câţiva ani mai târziu, Confederaţia generală a cadrelor franceze se pronunţă hotărât împotriva examenului psihotehnic.

Contra concluziilor categorice ale primilor testatori ai inteligenţei, celebrul ziarist Walther Lippmann susţinea, într-o dispută cu Lewis Terman (*apud* Gould, 1984, p. 210): „Fără a prezenta cea mai modestă dată asupra a ceea ce se petrece între concepţie şi vârsta de grădiniţă a copilului, ei anunţă, bazându-se pe ceea ce au extras din câteva mii de chestionare, că măsoară potenţialul mintal ereditar al fiinţelor umane... Dacă prevalează impresia că aceste teste măsoară inteligenţa, că ele constituie un fel de ultimă judecată asupra aptitudinilor copilului, că ele descoperă «ştiinţific» facultăţile sale predestinate, atunci este de o mie de ori mai bine ca toţi testatorii inteligenţei şi toate chestionarele lor să fie aruncate fără mustrări de conştiinţă în Marea Sargaselor".

Critica la adresa testului psihologic continuă şi în America, şi în Europa de Vest, dar atitudinea respectivă nu poate fi suspectată de rea-credinţă. Exagerarea calităţilor acestui instrument de investigaţie şi lipsurile, uneori severe, ale practicii profesionale justifică aprecierile acide faţă de test. În unele cazuri, criticii pătimesc prin exces. Pentru criticii cei mai înverşunaţi, ideea de a măsura faptele psihice este aberantă şi inumană. După unii dintre ei, testele nu au validitate reală, deoarece sunt disimulate, stupide şi fictive. Situaţia de examinare prin teste creează ea însăşi o tensiune mai mult sau mai puţin neliniştitoare, care determină ori influenţează grav rezultatele examinatului. Reacţiile la

teste sunt expresii ale situaţiei de testare. Despuierea rezultatelor şi diagnosticul sunt pline de arbitrar, după părerea acestor critici absoluţi, iar interpretările date de psiholog la elementele de răspuns ale candidatului traduc starea psihologului, şi nu starea psihică a examinatului.

1.4. Critici din partea unor organizaţii politice sau autorităţi guvernamentale

Desigur, ne întrebăm cum s-a ajuns la situaţia ca organizaţiile politice şi autorităţile statului să reacţioneze faţă de testul psihologic. Aceste reacţii sunt determinate de impactul social al teoriilor care stau la baza construcţiei unor teste şi de rezultatele şi concluziile la care se ajunge în urma aplicării testelor psihologice. Merită să amintim faptul că o serie de psihologi au fost militanţi sociali şi politici şi s-au implicat în pătimaşe dezbateri ideologice şi sociale. Reacţii pe această temă s-au înregistrat în fosta Uniune Sovietică, în Franţa şi în Statele Unite ale Americii. Urmările cele mai defavorabile pentru aplicarea testelor psihologice şi pentru psihologia aplicată s-au făcut simţite în fosta URSS.

În Uniunea Sovietică, testele sunt considerate mai întâi instrumente aflate în serviciul revoluţiei, apoi arme din arsenalul duşmanilor săi. Liderii mişcării testelor din URSS, Spielrein şi Gillerstien, vor să participe la transformările sociale în curs şi să constituie o psihologie marxistă. Ei explică diferenţele individuale de inteligenţă prin factori ereditari şi ambientali. Colegii lor occidentali le reproşau un ambientalism excesiv. Marile necazuri pentru psihologia din această ţară încep în deceniul patru al secolului XX, după ce, în 1931, mişcarea testelor atinsese apogeul, la vremea organizării la Moscova a celei de-a VII-a Conferinţă internaţională de psihologie aplicată (psihotehnică). Maurice Reuchlin (1967, p. 54) consideră că această conferinţă a fost o ocazie pentru delegaţii occidentali de a admira unele realizări importante în domeniul psihologiei diferenţiale, cum a fost Serviciul de selecţie profesională al Căilor Ferate, al cărui laborator central număra 80 de persoane (psihotehnicieni, medici, statisticieni). Acest laborator mai avea în jurisdicţie 20 de laboratoare în restul fostei URSS şi un vagon-laborator. Dar, în locul explicaţiilor de specialitate, oaspeţii au avut parte de lungi discursuri care proclamau necesitatea fondării ştiinţei pe principiile materialismului dialectic. Pe baza acestor principii, Lewitov aduce critici înverşunate la adresa testelor psihologice, care sunt adaptate, chipurile, pentru elevii provenind din mediul burghez şi care tind să prezinte ca imuabile particularităţile individuale datorate, în fapt, acţiunii mediului. Totuşi, chiar din 1929, începeau să apară critici la adresa testelor. În 1935, ca urmare a plângerii indignate a unui muncitor, Spielrein este judecat şi condamnat pentru că ar fi propus, într-un test cu răspuns multiplu, un răspuns considerat ofensator la adresa lui Stalin.

Criticile devin tot mai acerbe după 1931, polarizându-se asupra chestiunii diferenţelor individuale, pentru ca, în anul 1936, testele să fie blamate în mod oficial. Un decret al Comitetului Central interzice practica pedologiei, disciplină care, din 1930, se preocupa de copilul mic şi de tineret, practicând frecvent aplicarea testelor şi considerând puţin modificabile diferenţele observate. O rezoluţie a CC al PCUS „asupra deviaţiilor pedologice în Ministerul Educaţiei Publice" interzice pur şi simplu folosirea testelor psihologice. Sunt date două tipuri de justificări: primele cuprind ceea ce s-ar putea numi efectele perverse ale testelor în sistemul educativ. Practicii testelor i s-a atribuit

creşterea numărului de clase speciale. Pedologilor (practic, specialiştilor în psihologia copilului) li s-a reproşat că au subminat învăţătorii în procesul educativ şi, implicit, orientarea şi selecţia şcolară, definirea programelor, crearea de clase şi a unui învăţământ nou. Justificările principale sunt însă de ordin ideologic. Studiul şi practica testelor sunt judecate ca fiind incompatibile cu marxismul (ca şi genetica mendeliană). Pedologia are „tendinţe antimarxiste", „legea de subordonare a dezvoltării copilului la factori biologici şi sociali este în contradicţie absolută cu marximul", pedologia este „burgheză şi antiştiinţifică" şi cu atât mai periculos este faptul că „ea se disimulează sub o frazeologie marxistă". Pentru susţinătorii ideii potrivit căreia diferenţele individuale se explică 100% prin mediu, mediul este redus, de fapt, la influenţa uniformizatoare a ideologiei Partidului Comunist şi a învăţămintelor pe care el le propagă. În loc să studieze originea diferenţelor individuale, pedologia ar fi trebuit, într-o asemenea optică ineptă, să ajute la crearea unui „om nou", adică a unui robot uman ideologizat. Rezoluţia Comitetului Central a determinat suprimarea legislaţiei privind organizarea activităţii psihologului şi abolirea sistemului testelor, cărţile cu subiect pedologic fiind distruse. Mişcarea testelor psihologice dispare din rândul preocupărilor ştiinţifice. De fapt, este un segment profesional lichidat. Celelalte curente ştiinţifice, altele decât pavlovismul, au şi ele de suferit, dar nu dispar total.

Imediat după cel de-al doilea război mondial, atitudinea oficială de respingere a testelor psihologice se răspândeşte în partea de est a Cortinei de fier.

Spre anii '70, testele reîncep să apară în practica profesională şi ştiinţifică. În 1980, revista *Soviet Education* publică un număr special despre „reînnoirea testării şi măsura în şcoala sovietică". Decizia din 1936 a fost explicată prin abuzul practicii testelor, dar nu putea fi justificată interzicerea brutală a oricărei reflecţii şi a oricărei cercetări în acest sector. La ora actuală, numărul redus de psihologi cu specializare şi experienţă psihometrică este singurul factor care limitează practica testelor în ţările fostei URSS.

În Franţa, curentul psihologiei aplicate, care se manifesta prin practica testelor, s-a aflat în strânsă legătură cu mişcarea de reformă socială. Alfred Binet era preocupat de aplicaţiile psihologiei, dar transformarea societăţii nu reprezenta problema sa principală. În aceeaşi epocă, psihiatrul Édouard Toulouse se interesează de testele de inteligenţă, iar pozitivismul său îl orientează înspre chestiunile sociale. Interesul lui Toulouse pentru inteligenţă era în primul rând de ordin teoretic, dar el încearcă să precizeze relaţiile dintre geniu şi nebunie şi, pentru a îmbunătăţi observaţiile făcute, studiază persoanele recunoscute ca genii (Zola, mai ales), atunci când elaborează testele psihologice atât de celebre mai târziu. El întrevede utilizarea testelor în cadrul unei reforme a procesului de alocare socială a resurselor umane. Văzând în psihotehnică posibilitatea de a situa individul în funcţie de aptitudinile sale, şi nu în funcţie de privilegiul naşterii, Toulouse o concepe deci ca pe un factor de justiţie socială. Elevii lui, ca J.M. Jahy, H. Piéron sau H. Laugier, sunt angajaţi politic în mişcarea de stânga şi vor fi, în Franţa, liderii mişcării testelor, punând bazele psihologiei industriale şi ale orientării profesionale.

Cei care promovează testele sunt puternic angajaţi în acţiuni, pentru a reforma învăţământul. În 1905, Édouard Toulouse denunţa, deja, situaţia inechitabilă existentă în învăţământul francez, care favoriza accesul la treptele superioare de învăţământ doar a celor cu o solidă bază materială, şi, mai ales după primul război mondial, apare un puternic curent în favoarea unui tip de şcoală care să nu diferenţieze elevii din punct de vedere social. El este influenţat de către militanţi socialişti, sindicalişti şi de intelectualii

grupaţi în mişcarea *Les Compagnons de l'Université Nouvelle*. Printre aceşti intelectuali s-au numărat promotorii mişcării testelor. Ei acţionează pentru a crea un învăţământ de bază, până la 14 ani, comun pentru toţi, la încheierea căruia va fi operată o „justă selecţie". Pentru aceasta, se vizează introducerea testelor în examenul care trebuie să specifice concluziile după încheierea învăţământului de bază. Autorii proiectului nu ignoră faptul că, în medie, copiii burghezi reuşesc mai bine testele decât cei provenind din clasele populare. Acest aspect a fost observat „cu surpriză" de către Binet, apoi de către Stern, în 1914. Fără a exclude rolul eredităţii, inclusiv pentru a explica diferenţele între clasele sociale (lamarckiştii din psihologie considerau că nişte caractere achiziţionate pot deveni ereditare), ei sunt de părere că diferenţele individuale de inteligenţă sunt puternic determinate de condiţiile materiale (ca alimentaţia sau igiena) şi de educaţia familială şi şcolară. Laugier va conduce, în anii '30, mai multe anchete arătând legăturile dintre dezvoltarea biologică şi psihologică şi condiţiile de viaţă ale copilului. Nu este vorba pur şi simplu despre reformarea învăţământului, mai trebuie reduse şi inegalităţile sociale pe care le relevă psihotehnica.

În Franţa, cele mai aspre critici la adresa testelor au fost aduse de psihologii şi intelectualii membri ai Partidului Comunist, care găsesc perfect justificate măsurile luate în Uniunea Sovietică, în 1936. Referinţa la marxism şi voinţa afişată de a construi o „psihologie nouă" dau acestei critici o alură doctrinară, care nu se regăseşte în lucrările americane de aceeaşi tendinţă. Temele noii psihologii sunt expuse într-un editorial din revista *La Raison* (nr. 4 pe 1952, *apud* Huteau şi Lautrey, 1995). Dacă testele sunt criticate pentru caracteristicile proprii (baze teoretice slabe, artificialismul situaţiei de observaţie), atacul este şi mai înverşunat în privinţa funcţiei care li se atribuie. Testelor li se reproşează „că sunt mijlocul de a confirma, de a legitima şi chiar de a face să se accepte inegalităţile de ordin social". Autorii articolelor din revistă reamintesc că, între cele două războaie mondiale, mişcarea testelor a fost animată de anumite personalităţi progresiste şi că testele erau percepute ca mijloace de a modifica ordinea socială. Redactorii editorialului din *La Raison* amintesc de „opoziţia pasionată ridicată în mediile burgheze la primele aplicaţii ale psihometriei" (*apud ibidem*, p. 57). Dar ei consideră că partizanii şi adversarii testelor au fost victimele unei necunoaşteri fundamentale : testele, construite de către psihologi burghezi, nu pot fi decât burgheze, şi deci nu pot servi decât intereselor burgheziei. Concret, reprezentanţii acestei mişcări critice care denunţă anumite practici de selecţie a personalului (testele de personalitate erau, în fapt, mai vizate decât testele de inteligenţă) îşi propun să fondeze o didactică a informaţiei asupra meseriilor, care ar trebui să înlocuiască diagnosticul aptitudinilor. Mai târziu, ei vor duce o campanie împotriva „ideologiei darului". La sfârşitul anilor '60 şi începutul anilor '70, în cadrul mişcării de contestare, aceleaşi critici au fost din nou dezvoltate. În 1974, apare şi versiunea lor maoistă.

În Statele Unite ale Americii, a existat, cu siguranţă, tendinţa de a exagera calităţile testului psihologic, mai ales în domeniul studiului inteligenţei. Testele de inteligenţă au fost utilizate multă vreme pe continentul american, având ca scop justificarea superiorităţii unor grupuri rasiale asupra altora. Această stare de fapt l-a determinat pe David Hothersall să-şi intituleze un capitol din *History of Psychology* (1983) „Historical Uses and Abuses of Intelligence Testing" (pp. 300-330). Vizaţi în acest articol sunt, în principal, Lewis Terman, Robert Yerkes şi Henry Goddard, care au implicat rezultatele testărilor psihologice în patimaşe dispute ideologice şi rasiale.

Rezultatele acestor lucrări au fost utilizate pentru a recomanda practica segregației rasiale și sociale. Nivelul mintal al populației albe a fost apreciat ca fiind redus: vârstă mintală medie de 13 ani și 8 luni. Aceste „fapte" vor deveni argumente pentru eugeniști, care propovăduiesc restricțiile cu privire la viața sexuală a debililor, chiar sterilizarea. S-a constatat, de asemenea, că era posibil ca grupurile de emigranți să fie ierarhizate după rezultatele lor la teste: în frunte se aflau englezii, după care veneau în ordine alți *nordici* (germani, scandinavi, olandezi), apoi slavii, latinii și evreii. Datele respective au făcut obiectul dezbaterilor naționale asupra imigrării pe baza tezelor recomandând anumite cote naționale de imigrație. Psihologul C.C. Brigham a fost deosebit de activ în aceste dezbateri. Referitor la așa-numita inferioritate a negrilor (vârstă mintală 10 ani și 4 luni), concluziile se conformează ideilor segregației rasiale. Eugenismul, xenofobia, rasismul existau cu mult înainte de introducerea testelor de inteligență și este foarte dificil să se afle dacă exploatarea rezultatelor testelor din armată a avut sau nu un efect asupra deciziilor politice, ca aceea care a stabilit, în 1924, condiții de imigrare și mai restrictive. La fel de adevărat este că, în contextul societății americane a anilor '20, prin rezultatele testelor de inteligență (pentru că aceste teste pretindeau a măsura o inteligență generală ereditară) s-a alimentat un discurs reacționar.

Dacă majoritatea psihologilor americani erau ereditariști convinși (Goddard, Terman, Brigham vor abandona aceste convingeri în anii '30), în schimb anumiți lideri de opinie din Statele Unite nu împărtășeau aceeași credință, de unde o serie de controverse. Cea mai acerbă a fost aceea care opunea, în anii 1922-1923, pe Terman și Lippman, un elev al lui Yerkes, devenit jurnalist de notorietate mondială. Lippman nu lua în discuție calitatea testelor de inteligență și considera că ele puteau să aducă mari servicii, însă critica serios, și pe bună dreptate, pragul ales pentru definirea debilității și, mai ales, susținea ideea potrivit căreia nu era posibil să se evalueze inteligența independent de învățare, manifestându-se astfel contra ereditariștilor. El vedea în insuficiența mintală un efect reversibil al carențelor educative. Această controversă, care a avut puține efecte imediate, anunță dezbaterile și polemicile din anii '60.

În Statele Unite ale Americii, conflictele în jurul testelor devin sociale, chiar politice, puterea publică socotind inechitabilă folosirea testelor de inteligență pentru grupuri defavorizate. Parte din criticile anilor '50 au fost îndreptate contra abuzului de teste. De altfel, încă din 1945, produsele practicii psihologice industriale și educaționale intră în atenția publicului, mai ales prin testările de grup. American Psychological Association a avut grijă, în urma unor astfel de semnalări, ca testările de grup să nu ducă la abuzuri, care, la rândul lor, să determine măsuri guvernamentale împotriva folosirii testelor de grup.

Se pot enumera o serie de reacții din partea puterii legislative, executive și judecătorești. În 1964, senatorii Sam Erwin și Cornelius Gallager și-au unit eforturile pentru a interzice folosirea MMPI și a testelor nevalidate pentru angajarea de personal federal, cu excepția FBI. Președintele Lyndon Johnson a semnat o lege intitulată „Legea șanselor egale pentru toți de a fi angajați". O Comisie a șanselor egale pentru toți de a fi angajați (EEOG) pune legea în aplicare în 1964, prevăzând responsabilitatea de a ține sub control, în mod regulat, testarea psihologică. O serie de state, printre care New York și Texas, au luat atitudine cu privire la reglementarea folosirii testelor administrate individual. Domeniul imixtiunii legislative în psihologia aplicată a fost reactualizat în 1967, cu așa-numitul „caz Hobran Mansen". Judecătorul J. Skelley Wright a interzis folosirea procedurilor de testare psihologică în clase diferențiate, în funcție de capacități, în școlile din Washington, DC. În 1977, în cazul Larry P.V. Wilson Reles din California,

în care American Psychological Association a depus mărturie, ca și unii dintre cei mai importanți psihologi ai țării, judecătorul Robert Rachehan a interzis categoric folosirea testelor de inteligență standardizate (cum ar fi Wechsler sau Stanford-Binet scalar) în plasarea elevilor negri ori de alte naționalități în clase pentru educarea întârziaților mintal din statul California.

Prin aceste proceduri, cetățenii și reprezentanții federali au demonstrat că au capacitatea și voința de a interveni să reglementeze modul în care psihologii profesioniști își oferă serviciile și pot afecta viitorul potențial al clientului.

Joseph Mattarazzo apreciază că poziția psihologiei s-a înrăutățit din cauza standardelor inadecvate pentru interpretarea computerizată.

La originea conflictelor legate de testul psihologic au stat concepțiile asupra diferențelor individuale, sensibilitatea deosebită a unor pături sociale și rasiale la dezavantajarea lor prin examinărilor psihologice în care se folosesc teste. În Statele Unite, dezbaterile privind fundamentul diferențelor individuale în materie de inteligență generală au apărut la sfârșitul secolului al XIX-lea și nu au încetat, practic, niciodată. Disputele au depășit mediile academice, fiind o constantă a interesului public. Pentru mediile științifice, atenția s-a orientat mult timp asupra estimărilor relației dintre dobândit și moștenit, exprimată în termeni cantitativi, asupra fracțiunii de varianță fenotipică ce poate fi atribuită factorilor genetici într-o populație. *Ereditariștii* estimează această fracțiune la aproximativ 80%, *ambientaliștii* la mult mai puțin. Începând cu anii '70, o dată cu apariția revistei *Behavior Genetics*, crește priza de conștiință a marii relativități și ambiguități a măsurilor eritabilității (moștenitului, zestrei genetice) și se pune accent pe importanța interacțiunilor dintre factorii genetici și cei ambientali. Dar opozițiile nu se atenuează cu nimic. Cea mai vie polemică din perioada de după război demarează în anul 1969, când A.R. Jensen, profesor la Berkeley University, publică în *Harvard Educational Review* un lung articol de 120 de pagini, întrebându-se „până când vom putea să mărim QI-ul și reușita școlară". El prezenta, mai întâi, rezultatele care evidențiază eșecul vastelor programe de educație compensatorie destinate a facilita școlarizarea și integrarea copiilor din grupurile etnice minoritare. În al doilea rând, arată că în populația albă eficiența intelectuală și cea școlară erau puternic influențate de ereditate. În final, el declară că era rezonabil să se considere că factorii genetici joacă un rol important în explicarea diferențelor notabile ale capacității conceptuale dintre albi și negri (diferitele capacități asociative fiind, de asemenea, mai slabe). Dacă A.R. Jensen are câțiva susținători (R.J. Herrnstein, în 1971, și H.J. Eysenck, în 1973), majoritatea celor care participă la polemică socotesc inacceptabile argumentele ereditariste, autorul fiind acuzat de rasism, ceea ce nu rezolva însă problemele științei. Jensen devine unul dintre autorii cei mai citați negativ. Această dezbatere a permis estimarea audienței tezelor ereditariste, atât la psihologi, cât și în mass-media, care se dovedește mult mai slabă decât în anii '20-'30. Dezbaterile ideologice, din perioada interbelică, despre implicațiile rasiste ale concluziilor testelor psihologice ne informează mai curând despre autorii lor și despre conflictele ideologice ale societății americane în care se derulează acestea, decât despre testele în sine. În principal, criticile și dezbaterile sunt polarizate asupra chestiunii deformărilor sistematice de natură socială ale rezultatelor testelor (*biais* social) și asupra acelora privind fundamentul diferențelor dintre grupuri.

În 1994, apare o lucrare voluminoasă intitulată *The bell curve, intelligence and class structure in American Life*, publicată de psihologul R.J. Herrnstein și de un specialist în științe politice, C. Murray, și care a relansat această eternă dezbatere. În concepția celor

doi autori, societatea americană este din ce în ce mai stratificată, având drept bază eficiența cognitivă. Elita, care contribuie cel mai mult la progresul social și la producția de bogăție, este „cognitivă". Cu câteva excepții, săracii, șomerii, cei care au beneficiat de puțină școlarizare, delincvenții, părinții și cetățenii „răi" au un QI slab. Or, se întâmplă că șomerii sunt mai ales negri și că există în medie 15 puncte distanță între negri și albi. Herrnstein și Murray consideră că această constatare poate fi în mare măsură atribuită eredității. Deoarece nu este posibil să se ridice nivelul de inteligență al tuturor, pare de dorit, pentru binele colectivității, să se înceteze irosirea mijloacelor disponibile ale societății, ajutându-i în mod prioritar pe cei pentru care investiția socială este rentabilă, adică pe cei cu un QI considerabil. După un ritual acum bine cunoscut, această lucrare a suscitat în revistele psihologice specializate și în presa de mare tiraj numeroase comentarii, în general negative. Criticile cele mai frecvente nu au vizat testele ca instrumente de observare, ci explicația diferenței dintre albi și negri prin factorii ereditari. Poziția apărată de către Herrnstein și Murray este considerată ca derivând din îmbinarea unei ideologii conservatoare cu una rasistă.

Este adevărat că dispute au avut loc și în alte zone ale globului, dar dezbaterile au fost deosebit de animate mai ales în Statele Unite, unde corespund unor mize sociale puternice: modalitățile de selecție socială, mai ales școlară, și semnificația acordată diferențelor dintre albi, pe de o parte, și negri și hispanici, pe de altă parte. Aceste dezbateri au fost importate în Franța, de către câteva ziare mari și au luat alura dezbaterilor științifico-ideologice generale. Critici privind originea diferențelor individuale s-au emis și în URSS, dar, în ambianța politică totalitaristă, nu s-a permis manifestarea lor prin presă, așa cum s-a întâmplat în societățile democratice.

În Statele Unite, criticile în legătură cu avantajarea arbitrară a subiecților din clasele sociale superioare s-au dezvoltat ca urmare a publicării legii asupra drepturilor civile, lege al cărei obiectiv este eliminarea tuturor formelor de discriminare. Criticile aduse au avut două consecințe: în primul rând, o consecință juridico-administrativă (instituțiile judiciare și academice americane au impus anumite restricții practicii testelor) și apoi o consecință metodologică (psihologii americani au definit reguli mai stricte pentru construcția și aplicarea testelor, fiind grijulii să-și apere propriile tehnici).

În 1970, autoritățile din California constată o suprareprezentare a copiilor negri în clasele pentru retardați (deși reprezintă 10% din populația școlară, 25% din efectivele claselor pentru retardați sunt copii negri) și decid să suspende practica testelor de grup în școli. Aceleași autorități încurajează construcția testelor destinate capacității de a învăța, de a citi, teste care trebuiau să se substituie celor clasice. Era de așteptat ca noile teste să diferențieze copiii mai puțin în funcție de originea lor etnică. Este dificilă estimarea efectelor unei astfel de măsuri pentru că, în același timp, au fost definite cote indicând proporțiile de copii negri de nedepășit în școlile pentru retardați. Măsuri de același tip au fost luate în multe regiuni americane (Minneapolis, New York, Philadelphia). În alte zone (Chicago, mai ales) s-a recomandat utilizarea testelor în relație cu alte metode (observația comportamentală, observațiile clinice). Folosirea testelor în recrutarea profesională a fost și ea criticată, ajungându-se chiar în instanță; totuși, în selecția profesională s-au impus mult mai puține restricții.

Aceste atacuri au influențat psihologii și editorii de teste americani, în privința rigurozității construcției instrumentelor pe care le propun. Normele și standardele definite de către organizațiile profesionale ale psihologilor au devenit, de asemenea, mult mai precise.

A fost dezvoltată o metodologie pentru studiul distorsionării sistematice a răspunsurilor. Datorită acestei mase considerabile de lucrări realizate în chestiunea distorsionării sistematice, testele actuale și calitatea lor au evoluat destul de mult, diferențierea stabilită între grupurile sociale nemaifiind considerată arbitrară. Această diferențiere nu este, în general, mai puternică decât aceea constatată prin alte metode de investigație (diferențierile sociale arbitrare sunt mai pronunțate în acordarea notelor școlare sau la evaluările profesionale tradiționale decât la testele psihologice). Problema distorsionării de natură socială a rezultatelor examinărilor depășește domeniul testelor. Ea pune problema alegerii modalităților și a criteriilor de selecție socială, care se exersează, evident, mai mult sau mai puțin mascat, cu sau fără teste. Folosirea testelor a permis explicitarea și formalizarea acestei probleme.

1.5. Critici din partea beneficiarilor competenței psihologului

Criticile vin și din partea subiecților examinați, care nu au fost satisfăcuți de verdictul final al examinării și care au considerat că rezultatul le-a afectat situația profesională și prestigiul social.

Reacția din partea beneficiarilor competenței psihologului este explicată prin impactul pe care-l are testul asupra statutului și imaginii sociale și profesionale ale unor persoane, în urma comunicării unui rezultat al testării, nu neapărat favorabil. Nu ne putem aștepta ca o persoană respinsă la un examen de selecție profesională în care s-au folosit teste să se bucure de acest rezultat. Este bine cunoscut faptul că selecționerii din domeniul industrial sau din alte activități nu erau simpatizați de către cei examinați, mai ales în primele activități de acest gen, de la începutul secolului XX. Claude Lévy-Leboyer (1974, pp. 13-14) spune, în legătură cu organizarea muncii și selecția psihologică din mediul industrial: „*Cronometrorii* și *selecționerii* sunt rău văzuți. Selecția psihologică, la oricare nivel psihologic s-ar situa, este adesea interpretată ca o încălcare a personalității individuale. *Testatorii* sunt acuzați că folosesc metode fanteziste, în care validitatea nu este decât aparentă; ei condamnă anumiți indivizi la funcții subalterne, pe alții la șomaj, fără a se îngriji de consecințele diagnosticului lor; pe scurt, în general, sunt acuzați că împiedică jocul liber al alegerii profesionale". O asemenea optică, evidențiată, dar nesusținută în lucrarea citată, este negativistă, deoarece sunt omise avantajele nete ale intervențiilor psihologului, care constau în prevenirea accidentelor de muncă, evitarea eșecurilor profesionale și consilierea asupra potențialului persoanelor care participă la o selecție. Astfel de reacții negativiste sunt atitudini afective și merită să fie luate în seamă de metodologii examenelor psihologice în care se folosesc teste. Frecvența crescută a unor asemenea optici poate determina reacții legislative negative pentru psihologi. Ele pot conduce la îmbunătățirea unei metodologii, care trebuie să cuprindă intervenția unui ansamblu de metode în realizarea examenului psihologic. Reacțiile negative au venit mai ales din partea beneficiarilor testării psihologice, care și-au văzut imaginea socială influențată (în unele cazuri) în mod indezirabil.

Atrage atenția, în acest context, așa-numitul curent *open testing*, a cărui esență constă în atitudinea deschisă față de subiect, de comunicare convingătoare cu acesta, atitudine care să-l facă să accepte psihodiagnosticul ca pe un demers de determinare rațională și adecvată intereselor lui. Conform principiilor acestui curent, persoana căreia i se aplică

un test trebuie să fie bine informată cu privire la uzul ce se va da rezultatelor examenului la care este supusă. Claude Lévy-Leboyer susţine că psihologii americani au mers mai departe decât cei europeni şi supun frecvent dezbaterilor problemele legate de transparenţa rezultatelor testelor pentru cei care beneficiază de competenţa psihologului din domeniul selecţiei. În cadrul acestei noi optici a relaţiei profesionale, subiecţii trebuie să primească, înaintea examenului psihologic, o descriere a domeniilor cuprinse în testele pe care vor să le treacă, precum şi un eşantion de itemi. Mai mult, trebuie să le fie asigurată confidenţialitatea rezultatelor obţinute, ca şi liberul acces la oricare dintre testele care le-au fost administrate, după ce au fost corectate. În Statele Unite au fost cazuri în care psihologilor şi instituţiilor la care activau aceştia le-au fost intentate procese de către candidaţii nemulţumiţi de rezultatul examinării (care apreciau că alegerea testelor utilizate discriminează abuziv în privinţa persoanei lor). Atât psihologii, cât şi instituţiile au fost puse în postura de a demonstra pertinenţa testelor folosite. Lévy-Leboyer (1995, pp. 474-475) afirmă că astfel de conflicte nu au existat niciodată în Franţa, de exemplu, dar este de acord că trebuie respectat dreptul la confidenţialitate al persoanei examinate. O altă exigenţă, care decurge din practica *open testing*, este dreptul fiecăruia de a-şi cunoaşte propriile rezultate. Lévy-Leboyer (*ibidem*) spune: „Chiar dacă în principiu toţi psihologii sunt de acord în privinţa acestei exigenţe, mulţi se retrag în faţa dificultăţii pe care o presupune punerea ei în practică. În primul rând, rezultatele testelor sunt adesea exprimate sub formă de decile, de centile, de profiluri sau de note standard, care prezintă un caracter ermetic pentru neiniţiaţi. Mai grav, faptul de a comunica rezultatele testelor îi poate provoca subiectului examinat reacţii cu atât mai puternice cu cât este vorba de a-i pune în faţă handicapurile şi limitele sale". Psihologul nu trebuie, cu siguranţă, să evite obligaţia de a-i comunica subiectului rezultatele testelor doar pentru că aceasta este dificil de asumat, dar el trebuie să ştie să o facă în mod constructiv, în cadrul unui demers de consiliere, altfel spus, să nu furnizeze informaţii decât ca răspuns la întrebările puse de candidat şi să ajute subiectul să interpreteze datele, integrându-le în experienţa sa anterioară şi asumându-şi consecinţele lor.

Un exemplu în care competenţa psihologului este evidenţiată prin forme adecvate de comunicare îl oferă Laboratoarele de psihologie ale Căilor Ferate Olandeze (vezi Schonman, 1989, pp. 263-270). În laboratoare se acordă o importanţă deosebită testelor de siguranţă a circulaţiei şi proiectelor de automatizare lansate pentru a sprijini testarea psihologică. Rezultatul examinării psihologice are statutul unui sfat care, în practică, este foarte rar ignorat. Responsabilitatea pentru angajare revine conducerii locale, iar psihologul răspunde pentru sfatul dat. La Căile Ferate Olandeze nu se întocmesc rapoarte scrise. Psihologii le formulează oral şi fac doar notaţii pe dosar. Relaţia dintre conducerea locală şi psiholog este cea dintre client şi consultant. Consultantul este expertul, dar clientul este liber sau nu să urmeze sfatul dat. Solicitanţii au dreptul la o explicaţie în legătură cu rezultatul testării. Ei sunt liberi să permită sau să nu permită raportarea rezultatului către conducerea Căilor Ferate Olandeze.

În unele cazuri extreme ale conflictului psiholog – beneficiar, se ajunge în instanţa de judecată şi atunci proba psihologică este analizată, cu toată severitatea de către experţi. Astfel de exemple de punere sub acuzare a psihologului sunt relatate de Joseph Mattarazzo, Siro Sporli, Carl Hoyos şi Robert Gregory. Existenţa unor asemenea situaţii este un motiv foarte serios pentru a adapta proba psihologică la exigenţele curentului *open testing*. O soluţie pentru adaptarea testului psihologului la noile exigenţe poate fi

computerizarea, care să confere testelor un plus informaţional şi un suport pentru integrarea rezultatelor testului într-un ansamblu de metode psihodiagnostice.

Am evidenţiat şi exemplificat conţinutul reacţiilor antitest. Majoritatea reproşurilor sunt întemeiate şi trebuie mult efort pentru a ajunge la un remediu eficient. Constructorii testelor trebuie să ţină seama în activitatea lor de aspectele neconvenabile. Unele aspecte relevate nu ţin strict de metoda testului, ci de examenul psihologic, în general.

2. Dezbateri asupra calităţii testului psihologic în România

O dezbatere competentă asupra calităţilor şi defectelor testului psihologic a avut loc la 2 octombrie 1970, în cadrul unui simpozion intitulat *Valoarea şi limitele testelor psihologice*. Simpozionul, organizat de fosta Academie de Studii Sociale şi Politice, a beneficiat de participarea a 72 de specialişti din domeniile psihologiei generale, psihologiei pedagogice, psihologiei muncii şi psihologiei clinice. Dintre numele cunoscute ale psihologiei româneşti, au participat C. Botez, A. Cosmovici, V. Ceauşu, R. Floru, D. Muster, P. Pufan, I. Radu, M. Roşca, Al. Roşca şi B. Zörgo. Motivaţia întâlnirii a fost expusă de Al. Roşca în trei puncte:

- cerinţa acută de teste psihologice şi numărul redus de teste valabile aflate la îndemâna psihologilor;
- numărul redus de specialişti, care pot aplica în mod riguros şi pot interpreta ştiinţific rezultatele;
- necesitatea unui schimb de experienţă între psihologii care aplică teste.

Cele mai importante idei şi propuneri au fost, prezentate punctual, următoarele:

- Testul psihologic constituie principalul instrument al psihologiei diferenţiate şi aplicate. Dar testele nu sunt decât nişte instrumente, iar competenţa celor care le aplică joacă un rol hotărâtor în obţinerea unor rezultate valabile (C. Botez, D. Vacaru, B. Zörgo, M. Roşca, Gh. Bontilă, R. Floru).
- Este de dorit ca, în cadrul examenului psihologic, testele psihometrice să se îmbine cu metoda clinică (observaţia, interviul, anamneza etc.). Se recomandă utilizarea cu discernământ a testelor şi chestionarelor de personalitate (A. Cosmovici, I. Vianu, M. Cristescu, M. Mamali).
- Elaborarea testelor trebuie să respecte riguros tehnicile statistice de etalonare şi validare, pentru ca acestea să devină instrumente valabile de lucru (I. Radu, M. Mamali, V. Ceauşu).
- Este nevoie să se aprofundeze metodele de validare, în sensul completării validării empirice cu cercetarea structurilor funcţionale implicate în realizarea sarcinilor (B. Zörgo, V. Ceauşu, S. Marcus, I. Radu, P. Pufan).
- Se impune ideea înfiinţării unei comisii metodologice şi a stabilirii de norme privind metodele de examinare psihologică, de elaborare, adaptare şi utilizare a testelor (A. Cosmovici, R. Floru, D. Voinescu, C. Botez, M. Bolos, D. Muster).
- A fost hotărâtă înfiinţarea unui organism de elaborare şi difuzare a unor metode concrete de examinare psihologică, îndrumare şi control al activităţii practice de

orientare şi selecţie profesională la Ministerul Muncii (C. Botez, D. Vacaru, D. Muster, I. Radu, F. Iacomi).

- S-a susţinut necesitatea organizării unor măsuri de specializare pentru pregătirea cadrelor în raport cu cerinţele mereu în creştere (D. Vacaru, C. Botez, R. Floru, Gh. Bontilă).
- Este necesară înfiinţarea unor centre, oficii sau laboratoare-model, care să se ocupe cu orientarea profesională a tinerilor (A. Cosmovici, F. Iacomi, D. Vacaru, C. Botez).
- Au fost propuse demersuri pentru îmbunătăţirea planurilor de învăţământ ale secţiilor de psihologie (A. Cosmovici).
- Au fost stabilite criterii de angajare a psihologilor în industrie (A. Cosmovici, P. Pufan, B. Zörgo).

Deşi au trecut mai bine de treizeci de ani de la formularea unor asemenea observaţii, se poate spune că ele îşi păstrează actualitatea.

Printre documentele ştiinţifice ale reuniunii figurează şi cinci studii critice asupra testului psihologic, semnate de C. Botez (*Utilizarea testelor pentru depistarea şi aprecierea aptitudinilor profesionale*), A. Cosmovici (*Aplicarea testelor pentru determinarea nivelului de dezvoltare intelectuală*), I. Radu (*Metoda testelor văzută din perspectiva statisticii psihologice*), M. Roşca (*Valoarea şi limitele metodelor de diagnosticare a personalităţii*), B. Zörgo (*Criteriile unui test ştiinţific*).

C. Botez consideră că s-a ajuns la compromiterea unor teste prin administrarea acestora de către persoane necalificate, prin vulgarizarea tehnicii şi întrebuinţarea necorespunzătoare. Tehnicile psihometrice au o precizie slabă. C. Botez îşi sprijină argumentaţia pe cazul semnalat de Thorndike, în care 1.000 de supradotaţi ai lui Terman dădeau alte rezultate peste un număr de ani, ceea ce nu mai justifica acordarea calităţii respective. De asemenea, aminteşte de abuzurile din anii '50, din SUA, în privinţa folosirii testelor. Invazia testelor în viaţa particulară a dus la proteste din partea corpului didactic, a organizaţiilor de părinţi şi chiar din partea psihologilor. Se dă, în acest ultim caz, exemplul profesorului Carl V. Smith de la University of Wisconsin, care neagă testelor orice valoare. Printre cauzele discreditării unor teste se numără concepţia privind aptitudinea şi măsurarea lor, precum şi modul de aplicare şi interpretare. Metodele de validare şi etalonare sunt insuficient cunoscute de către specialişti. Examinarea efectuată de specialişti trebuie combinată cu observarea comportamentului în timpul examinării. Autorul amintit consideră ca importante modul de comportare faţă de erori şi indicaţiile referitoare la temperament şi caracter. Completarea rezultatului trebuie făcută cu o fişă de observaţii din timpul probei, cu datele biografice, cu analiza produselor activităţii, cu aprecierile colegilor, profesorilor şi părinţilor subiectului şi caracterizarea de pe teren, folosind date obiective şi aprecieri personale.

Profesorul Andrei Cosmovici a abordat, în studiul amintit, doar testele pentru determinarea nivelului de dezvoltare intelectuală, cu referire la două categorii: testele care se folosesc pentru a determina dezvoltarea copilului şi testele care îşi propun să stabilească deosebiri de nivel intelectual între persoane de aceeaşi vârstă. În practică, se produce o confuzie între diferenţele de vârstă şi cele datorate nivelului aptitudinilor indiferent de ea. Testele trebuie să stabilească ce poate acum subiectul, dar să se refere şi la posibilităţile de progres. Autorul ieşean afirmă, de asemenea, că testele dau o indicaţie

incompletă, inexactă asupra aptitudinii generale. Factorii extraintelectuali au o influenţă directă în rezultatele testelor.

După părerea profesorului Andrei Cosmovici, inteligenţa generală nu se poate diagnostica printr-un singur tip de probă. Probele folosite pentru această diagnosticare trebuie să fie numeroase şi variate, nonverbale şi verbale. Rezultatele la lucrările de sinteză psihodiagnostică necesită trei ore de travaliu intelectual. Convorbirea şi observaţia în timpul probei trebuie să completeze rezultatele cantitative. De asemenea, se recomandă un examen individual şi interpretarea probelor utilizate de către personal specializat şi cu experienţă.

I. Radu insistă, în studiul său, asupra semnificaţiei valorilor numerice pe care le obţinem prin teste şi consideră că acestea nu au depăşit puterea informaţională dată de scalele ordinale. În funcţie de forma curbei de distribuţie, trebuie să folosim indicatori diferiţi în exprimările cantitative. Simplitatea în materie de valori numerice nu înseamnă exactitate din punct de vedere psihologic. În cazul răspunsului la 12 întrebări, punctajul egal cu 5 poate fi obţinut în sute de feluri. Profesorul clujean acordă atenţie necesităţii de a asigura o anumită consistenţă internă a testelor şi o notare mai nuanţată.

Mariana Roşca se referă cu prioritate la chestionarele de personalitate, tehnicile proiective şi testele obiective, atrăgând atenţia asupra situaţiilor de falsificare a răspunsurilor prin deformarea deliberată a lor şi prin mecanismele de apărare.

Beniamin Zörgo spune, la începutul studiului său, că verificarea şi perfecţionarea continuă a instrumentelor şi procedeelor utilizate în cunoaşterea personalităţii reprezintă sarcini permanente ale psihologiei. Criteriile ştiinţifice trebuie să funcţioneze nu doar în faza de elaborare, ci şi în faza procedurală, adică de la efectuare până la interpretarea rezultatelor. Aprecierea unei persoane trebuie să se facă pe baza unei analize cantitative şi calitative, folosindu-se cele mai avansate metode statistice (autorul nu le enumeră însă), şi pe baza interpretării a numeroase date, obţinute prin aplicarea unui mare număr de probe şi a mai multor metode de investigaţie. B. Zörgo aminteşte că abuzurile în folosirea testelor au dat naştere la proteste în ţările lumii civilizate. Greşelile şi abuzurile sunt comise, în primul rând, de diletanţi şi duc la compromiterea ştiinţei numite psihologie, la degradarea rezultatelor pe care le-a înregistrat psihologia ştiinţifică. În privinţa acestei aprecieri, considerăm că este vorba despre o exagerare: greşelile de diagnostic şi tratament ale medicilor nu au dus la compromiterea medicinei, de exemplu.

O contribuţie relativ recentă la analiza psihometriei clasice o aduce Stefan Szamosközi (1997, p. 187), care consideră că psihometria tradiţională este statică şi constatativă, deoarece indică rezultatul procesului cognitiv, şi nu oscilaţiile procesului intelectual individual: „Ca şi în cazul unui lanţ oarecare, dacă o singură verigă este slabă, tot lanţul este slab. Metodele clasice ne spun numai că lanţul este slab, nu ne spun care anume este veriga (verigile) care determină această deficienţă. Doi subiecţi cu un QI de 70 sunt amândoi consideraţi ca având intelect de limită. Mecanismul (sau mecanismele) care au produs această deficienţă cognitivă pot fi total diferite la unul dintre subiecţi faţă de celălalt. Oricum, psihometria clasică ne semnalează doar rezultatul funcţionării cognitive, nu şi procesualitatea ei" (*ibidem*, p. 188).

Aşadar, instrumentul clasic de evaluare a inteligenţei ne spune doar care este performanţa, nu şi cum o putem modifica. Prin urmare, diagnosticul clasic nu poate dezvolta intervenţia. Pentru a putea interveni în mod eficient, trebuie să cunoaştem componenta sau mecanismul cognitiv deficitar. Metodele dinamice de diagnosticare implică o altă

relaţie între examinat şi examinator şi antrenează modificarea testului însuşi care, de exemplu, măsoară câştigul cognitiv ce intervine între pretest şi posttest condiţionat de învăţare. În consecinţă, spune Stefan Szamosközi, diagnosticul devine şi formativ, pentru că el poate dirija intervenţia, modificare cognitivă necesară pentru ameliorarea performanţei intelectuale. În studiu, se fac referiri competente şi documentate la modificările apărute în examinarea psihologică pentru ca aceasta să corespundă exigenţelor evaluării formative, care încearcă să elibereze eficienţa intelectuală de barierele sale noncognitive, prin oferirea unui suport informaţional permanent, vizând activarea potenţialelor intelectuale proprii nivelului real al funcţionării cognitive.

Criticile aduse metodei clasice de examinare prin teste se referă la natura standardizată a testelor de inteligenţă, la examinarea de tip interogatoriu (în cursul căruia subiectul nu este informat asupra pertinenţei răspunsurilor sale), la neutralitatea examinatorului.

Capitolul VI

ASPECTE ALE PRACTICII TESTULUI PSIHOLOGIC

1. Domenii de aplicabilitate

Dacă o serie de discipline, mai ales exacte, au trebuit să aştepte o bună perioadă de timp pentru a-şi vedea modelele aplicate, psihologia a fost de la începuturi o disciplină aplicativă, în special prin crearea testelor. Uneori însă, lipsa de precizie a măsurătorilor prin intermediul testului psihologic, superficialitatea aplicaţiilor şi mai ales a concluziilor au dus la apariţia îndoielii cu privire la valabilitatea acestui instrument de investigaţie. Aplicaţiile psihologiei nu se reduc la domeniul testelor, dar într-o bună parte a domeniilor de aplicaţie intervine şi testul, rar ca metodă unică, deseori în combinaţie cu altele.

1.1. Domeniul psihopedagogic

Domeniul psihopedagogic a fost şi rămâne unul dintre cele mai vaste câmpuri de aplicaţie pentru testul psihologic, având în vedere caracterul de reţea permanentă, de o deosebită densitate, al instituţiei şcolare. Este motivul pentru care rezultatele sunt foarte numeroase şi servesc ca bază de etalonare pentru o bună parte dintre testele existente.

Primele „comenzi sociale" au apărut în domeniul psihopedagogiei speciale şi au constat în solicitarea detectării copiilor cu deficienţe intelectuale, în vederea creării unui învăţământ special. Binet şi Simon au onorat din plin această primă comandă socială şi au generat încredere în instrumentul de investigaţie pe care l-au lansat.

Aplicarea măsurătorilor caracteristicilor psihice prin teste în domeniul şcolar este frecventă atât în slujba activităţii de învăţare, cât şi în cea de cercetare, unde sunt evidenţiaţi factorii care condiţionează randamentul şcolar. În domeniul cercetării, testul este util şi pentru probarea eficacităţii diverselor metode de predare, pentru formarea grupurilor de randament, pentru optimizarea utilizării mijloacelor audiovizuale în prezentarea materialului de învăţat şi pentru verificarea efectelor instruirii programate şi automatizate. Pentru realizarea acestor deziderate avem nevoie, în primul rând, de un psihodiagnostic de bună calitate, adică stabilit în conformitate cu o metodologie bine pusă la punct şi adaptat nevoilor speciale ale învăţământului. Astfel, când şcolarii nu fac un progres corespunzător în lectură şi calcul, se poate evidenţia nuanţat, cu ajutorul testelor psihologice, natura dificultăţilor pe care le întâmpină. Se poate stabili dacă şcolarului îi lipseşte o aptitudine specială care îi este necesară pentru un progres normal sau dacă dificultăţile de învăţare sunt provocate printr-un anumit fel de dificultate senzorială.

Testele pot fi utilizate în domeniul consilierii intereselor profesionale şi în cel al aptitudinilor profesionale, atunci când la sfârşitul unui anumit ciclu educaţional se fac recomandări pentru un domeniu şcolar ulterior sau pentru o activitate profesională. Testele existente în prezent, pe plan mondial, sunt foarte numeroase şi diversificate, în funcţie de vârstă şi de stadiile de dezvoltare, fiind aplicate atât în stadiul preşcolar, cât şi în cel primar şi secundar. De obicei, testele nu sunt aplicate înainte de şcoala secundară. Ţările puternic dezvoltate în plan economic (etalonul reprezentându-l Statele Unite ale Americii) au pus la punct programe bine fundamentate în acest sens (ne referim în special la *American Testing Program*).

Un domeniu vast de aplicare a testelor are ca punct central adaptarea elevilor la complexul industrial şi social, în care munca trebuie să fie divizată şi organizată corespunzător, încât aptitudinile şi înzestrările speciale ale tinerilor, progresiv intraţi pe piaţa muncii, să fie valorificate. Probleme de acest fel trebuie să facă obiectul de interes pentru programe de cercetare de mare anvergură, care să fie implementate de cercetători deosebit de experimentaţi şi talentaţi. O bună parte dintre dificultăţile sociale şi economice ale unei ţări au la bază contraselecţia valorică realizată la nivelul tinerei generaţii, însoţită de un proces de formare deficitar şi, firesc, ineficient.

O direcţie de prim rang în aplicarea instrumentelor de măsurare psihologică este cea care se ocupă cu măsurarea progresului randamentului. Un mare număr dintre testele produse de edituri sunt teste de randament şcolar, create şi de profesori pentru aspectele educaţionale cu care se confruntă ei. Pentru a fi corect construite, aceste instrumente de investigaţie ar trebui să fie consecinţa colaborării dintre psihologi şi profesori cu mare experienţă de predare. Experienţele de până acum arată că este nevoie de o lungă perioadă de timp pentru a se putea elabora teste obiective de randament şcolar, realizare condiţionată de exigenţele tehnice de grad înalt ale măsurării psihologice şi de alte aspecte teoretice proprii disciplinelor de învăţământ. Dezvoltarea instrumentelor de măsură satisfăcătoare, care presupun o cantitate imensă de noţiuni teoretice şi material tehnic, nu poate fi pretinsă profesorilor şi psihologilor şcolari supraaglomeraţi şi subplătiţi.

Din punctul de vedere al relevanţei informaţionale a rezultatelor testelor, o atenţie deosebită trebuie acordată detectării surselor de influenţă diferenţiată asupra acestora şi elaborării diferenţiate a etaloanelor. Aplicarea testelor în domeniul şcolar, în ţările cu o compoziţie variată din punct de vedere rasial, a dus la apariţia unor probleme delicate în urma emiterii unor concluzii pripite. De exemplu, anumite concluzii raportate în privinţa nivelului de inteligenţă al copiilor negri, defavorizaţi clar mai ales la probele verbale, ar trebui explicate din perspectiva ambientală, prin condiţiile de viaţă caracterizate de izolare şi marginalizare.

O altă sursă de distorsionare a rezultatelor o constituie nefamiliarizarea cu activitatea de examinare prin intermediul testelor, fapt ce necesită, în majoritatea cazurilor de acest fel, explicaţii suplimentare pentru a determina dezinhibarea subiecţilor. Ne referim la acele instituţii şcolare situate în zone geografice pentru care testarea psihologică reprezintă o noutate. Timpul de solicitare trebuie să fie în aşa fel stabilit încât să nu provoace oboseală prematură copilului. Precauţii deosebite trebuie să fie luate şi în modul de furnizare a rezultatelor, ţinându-se cont de sensibilitatea copilului şi a părinţilor. Există posibilitatea ca rezultatele experienţelor să aibă consecinţe grave asupra subiecţilor. De exemplu, sunt diferenţe notabile între cazul elevilor care participă la experimentarea unui nou învăţământ şi cei care urmează un învăţământ tradiţional.

Cabinetele psihologice şcolare au în general o lungă tradiţie, mai ales în ţările dezvoltate. Familiarizarea copilului cu testul poate constitui, de asemenea, o sursă de deformare a rezultatelor. În instituţiile şcolare cu tradiţie este realizat un profil psihologic şcolar, profil care indică gradul de dezvoltare a unor procese psihice şi o analiză a ansamblului personalităţii.

Aplicarea testelor în domeniul şcolar favorizează verificarea şi asigurarea criteriilor de valabilitate a acestora (ne referim la fidelitate, validitatea concomitentă şi predictivă). Dintre tipurile de teste, în domeniul şcolar sunt foarte uzitate testele creion-hârtie. În cercetarea fidelităţii, perioadele de timp în care testul trebuie repetat sunt mult reduse, deoarece la trecerea unei perioade considerabile de timp avem de-a face cu „alţi" copii, care au evoluat în privinţa nivelului caracteristicilor studiate. Accelerarea intelectuală despre care se vorbeşte tot mai frecvent duce la uzura accentuată a unor teste, la neconcordanţa rezultatelor aplicate pe acelaşi nivel de vârstă. Aşadar, în analiza fidelităţii sunt de preferat metodele testului paralel, metoda înjumătăţirii şi metodele Kuder-Richardson.

1.2. Domeniul resurselor umane

Sub denumirea de *resurse umane* găsim preocupări vechi şi noi din diverse domenii de activitate. Dinamica deosebită a acestui sector teoretic şi practic al preocupărilor umane face ca partea cunoştinţelor recente să ia o pondere din ce în ce mai mare.

În privinţa specialiştilor practicieni, care lucrează în domeniul resurselor umane, se face distincţie între „generalişti", preocupaţi de ansamblul problemelor apărute în departamentele de specialitate, şi „experţi", adică persoane beneficiind de o pregătire specializată pentru un anumit sector, cum ar fi recrutarea sau perfecţionarea personalului, şi preocupate numai de activitatea acestuia. Experţii pot fi psihologi, economişti, jurişti, sociologi, pedagogi, asistenţi sociali, statisticieni, normatori etc. În practica departamentelor de resurse umane din România, „generaliştii" sunt titraţi universitari, de o mare varietate în privinţa profilului de studii. Statutul de „generalist" nu exclude posibilitatea ca persoana să fie specializată într-un anumit domeniu, dar, pe măsură ce creşte dimensiunea demersului, atribuţiile de coordonare şi de abordare interdisciplinară devin prioritare. Pe plan internaţional, atât „generaliştii", cât şi experţii fac parte din asociaţii profesionale şi sunt atestaţi de către acestea sau de institute de atestare pentru activitatea pe care o desfăşoară.

Ce atribuţii are un psiholog în cadrul domeniului de resurse umane? Principalele activităţi ale psihologului în cadrul unei organizaţii sunt recrutarea, promovarea şi prognosticul aptitudinilor şi capacităţilor profesionale. Prognosticul se realizează mai ales cu privire la competenţele, activitatea, adaptibilitatea persoanelor, la capacităţile acestora de a se integra în grup sau de a-l conduce. Un astfel de prognostic este util pentru întreprindere în termeni de randament, de performanţă şi de climat. În materie de resurse umane, „eşecurile costă foarte mult". Psihologul abordează probleme practice având, prin formaţia sa universitară, deprinderea unor metode şi un anumit mod de raţionament care vor rămâne fondul său profesional, dar în practică, în domeniul concret, demersul său este adesea empiric. Instrumentele pe care le utilizează psihologul trebuie să-i ofere informaţii mai mult sau mai puţin pertinente. Un psiholog se formează cu adevărat după ce lucrează o perioadă de timp în mediul pentru care s-a pregătit şi

realizează observaţii, convorbiri cu personalul angajat, lucrând, în unele cazuri, împreună cu acesta. Astfel, psihologul îşi va dezvolta o sensibilitate aparte şi va constata justeţea aprecierilor sale.

Din punctul de vedere al preocupărilor profesional psihologice, considerăm că acest domeniu înglobează o bună parte din sectorul clasic, denumit selecţie şi orientare profesională. Nu este bine în nici un caz să punem semnul de egalitate între aceste două sectoare de preocupări profesionale, deoarece în vremurile moderne activitatea psihologului din domeniul resurselor umane presupune interacţiunea crescută cu alţi specialişti, vizând optimizarea alocării persoanelor pentru diverse activităţi, prestate individual sau în cadrul organizaţiilor.

În acest domeniu, sunt utilizate o gamă diversificată de teste în care excelează testele de aptitudini, testele de cunoştinţe şi, nu în ultimul rând, testele de personalitate. Psihologul foloseşte testul ca instrument de psihodiagnostic, într-un ansamblu de metode psihodiagnostice : convorbirea, experimentul (mai rar), analiza biografică etc. Sursele noastre bibliografice ne indică o proporţie variabilă de folosire a testului.

În 1988-1989, s-a realizat o anchetă pe un eşantion de întreprinderi private, naţionalizate şi cabinete de recrutare franceze. Tabelul recenzează tehnicile de evaluare utilizate în mod obişnuit sau ocazional de diferite tipuri de organizaţii şi, în partea de jos, tehnicile utilizate pentru decizia finală (*apud* Cadin, Guérin şi Pigeyre, 1997, p. 186).

Tabelul 41

Frecvenţa utilizării instrumentelor de investigaţie psihologică în domeniul resurselor umane

Tehnici utilizate	Serviciul de recrutare			
	Cabinete (%)	Întreprinderi private (%)	Întreprinderi naţionalizate (%)	Total eşantion (%)
Convorbire	98	100	100	99
Examen grafologic aprofundat	97	96,5	69	93
Teste de aptitudini sau de inteligenţă	55	69	84,5	63
Teste de personalitate	61,5	55	84,5	63
Minisituaţii de muncă	28	41	46	34
Tehnici proiective	21,5	17	23	20
Alte tehnici (morfopsihologie, analiză tranzacţională)	25	0	0	15

Cifrele din tabel nu reflectă decât datele unui sondaj, dar evidenţiază o parte dintre tipurile de teste utilizate în acest domeniu. În cazul în care psihologul îşi practică profesiunea într-un cabinet propriu de consiliere în domeniul resurselor umane, exigenţele de calitate privind testele trec în prim-plan, deoarece, în cazul unor consilieri individualizate, va trebui să se găsească în testele psihologice un suport bine constituit pentru efectuarea de prognoze.

1.3. Domeniul transporturilor

Transporturile ca domeniu de intervenţie a psihologiei aplicate au o tradiţie respectabilă în realizări psihometrice.

Primul pas în acest sens îl face Hugo Münsterberg, în 1910, pe când activa la American Association of Labor Legislation, prin crearea unui test de aptitudini profesionale pentru vatmani. Acest test reproduce pe o foaie de hârtie, prezentată subiectului, situaţii puternic tipizate din circulaţie. Metoda dezvoltată pentru prima dată de Münsterberg prezintă primul procedeu de selecţie pentru participanţii la circulaţia rutieră. Ideile sale de construcţie a testelor au fost preluate în câteva ţări prinse în primul război mondial, în scopul selecţiei şoferilor de camioane.

Activitatea de creare de teste în domeniul transporturilor se derulează altfel în Germania, unde, în anul 1917, sunt elaborate probe noi destinate selecţiei personalului de exploatare a societăţilor de tramvaie şi a celui de la Căile Ferate. Pe lângă acestea, sunt create şi dispozitive de exerciţiu (forme de simulatoare) pentru selecţia şi formarea conducătorilor de mijloace de transport. În anul 1917, începe la Eisenbahn-Generaldirektion Dresden construirea unui laborator psihologic pentru mecanicii de locomotivă. De asemenea, la Berlin a fost înfiinţat, în ianuarie 1917, un laborator psihotehnic pentru vatmani. În Hamburg, William Stern şi colaboratorii săi au desfăşurat examene de aptitudini pentru cei ce optau pentru profesiunea de vatman, cu scopul declarat de a reduce numărul accidentelor. Până în primăvara anului 1920, Căile Ferate şi Poşta germană solicitau posturi corespunzătoare de psihotehnicieni. În timpul războaielor mondiale, activitatea de selecţie pentru domeniul transporturilor militare este deosebit de intensă nu numai în Germania, ci şi în majoritatea ţărilor implicate.

Actualmente, laboratoarele de examinare psihologică din transporturi, care utilizează masiv testele în activitatea lor, sunt bine organizate în majoritatea ţărilor din lume. O bună parte dintre aceste laboratoare sunt organizate în reţea, mai ales în domeniul transporturilor feroviare şi aeriene. Dotarea lor cuprindea o proporţie apreciabilă de teste aparative, deosebit de costisitoare. În activitatea de examinare din acest domeniu s-a făcut apel, în bună parte, la computerizare. Faptul că, în acest domeniu, examinarea este permanentă permite colectarea unui mare număr de date, ceea ce înlesneşte verificarea permanentă a valabilităţii testelor utilizate.

1.4. Psihologia clinică

În acest domeniu, se utilizează mai ales probe proiective şi chestionare de personalitate, dar nu este exclusă folosirea testelor de randament motrice, senzorial şi intelectual. Aplicarea chestionarelor şi testelor de personalitate este pusă în serviciul precizării diagnosticului şi al stabilirii terapiei adecvate (în cazul în care psihologul lucrează în mediul spitalicesc, în colaborare cu medicul psihiatru). O valoare complementară, fără dubii, o au cunoştinţele de psihologie clinică. O problemă care trenează în domeniul clinic este aceea a raporturilor dintre medicul psihiatru şi psihologul clinician.

Cunoscutul terapeut Michael Balint consideră (în Balint şi Balint, 1976) că psihologul, în calitate de colaborator al psihiatrului, are în clinica psihiatrică următoarele cinci roluri :

- pregătirea pacientului (în sensul unei convorbiri de adaptare la situaţia de examinare prin test) ;

- administrarea testelor;
- înregistrarea răspunsurilor pacientului;
- evaluarea răspunsurilor;
- validarea testelor.

Desigur, psihologii clinicieni au şi alte atribuţii, dar ne referim doar la domeniul aplicării testelor. Într-o bună parte dintre cazuri, psihologii clinicieni lucrează în mediul spitalicesc nepsihiatric şi atunci colaborează cu medicul curant pentru a adapta persoana aflată în suferinţă la noua sa condiţie.

Printre probele de personalitate cu o respectabilă tradiţie în aplicarea clinică se numără testele Rorschach, Szondi, Lüscher, Chestionarele 16 PF, proba Eysenck (intro/extraversiune şi nevrozism), Woodworth-Mathews, Fahrenberg-Selg. Trebuie să amintim faptul că a existat, şi există încă, o anumită opoziţie la folosirea metodelor cantitative şi a procedeelor de măsurare în domeniul clinic.

1.5. Domeniul psihologiei judiciare

Testul psihologic şi-a relevat potenţialul aplicativ în acest domeniu mai ales prin intermediul testelor de personalitate. Aplicarea lor pune psihologilor probleme serioase, atunci când sunt folosite în domeniul probaţiunii. În acest caz, o importanţă aparte capătă procedeele menite să pună în evidenţă erorile de măsurare datorate intervenţiilor unor factori sistematici.

1.6. Domeniul cercetării ştiinţifice

Testul psihologic este strâns legat de activitatea de cercetare ştiinţifică. Elaborarea testului psihologic poate fi considerată o activitate de cercetare ştiinţifică, prin numeroasele date de natură diferită care trebuie colectate, prin ipotezele elaborate, prin verificarea lor, prin realizarea unor demersuri experimentale. După elaborarea şi verificarea criteriilor de apreciere a testului, acesta poate fi utilizat în cercetarea ştiinţifică în calitate de instrument valid de investigaţie. Activitatea se poate desfăşura sub forma investigării unor eşantioane de subiecţi, cu respectarea regulilor metodologiei experimentale în care se constată nivelul unor aptitudini sau capacităţi, la un moment dat, înainte şi după introducerea unui tratament experimental.

Cercetarea ştiinţifică poate fi realizată longitudinal sau transversal şi este pusă mai ales în serviciul psihologiei diferenţiale. Probele transversale pot servi la compararea randamentelor diferitelor grupuri de subiecţi, pentru a stabili semnificaţia diferenţelor observate. Cercetarea ştiinţifică poate fi realizată în domeniul senzorial-motrice, cognitiv, afectiv-volitiv, psiho-social etc.

1.7. Domeniul psihologiei sportive

Profesioniştii psihologi care lucrează cu sportivii apelează la testul psihologic atunci când se pune problema conturării unui profil psihologic al jucătorilor de echipă sau al celor care practică atletismul. De la aceste date se pleacă pentru realizarea *training*-urilor psihologice. Fiind evident că sportul, în general, a devenit o afacere întinsă, interesează

în mod deosebit orice sursă psihică exploatabilă pentru îmbunătăţirea performanţelor sportive. Dar, pentru a fi exploatată, aceasta trebuie să fie mai întâi corect diagnosticată. Desigur, apar probleme de etică, presupunând o rezolvare adecvată, dar acestea constituie un aspect care trebuie tratat de cercetători cu preocupări specializate. În domeniul sportiv, se aplică atât probe de randament psihic, cât şi teste de personalitate.

2. Psihologul şi problemele deontologiei

Perioade îndelungate de timp, majoritatea psihologilor s-au arătat puţin interesaţi sau preocupaţi de problemele de etică. Presiunile practicii de cercetare şi cele ale activităţii de psihodiagnostic le-au impus să ia în considerare anumite standarde de natură etică. Acumularea unor disfuncţionalităţi în practicarea profesiei i-a determinat pe psihologi să le ceară guvernanţilor elaborarea unor prevederi legale privind activitatea de psihodiagnostic. Adoptarea unor prevederi legale este o realitate în ţările dezvoltate în plan socioeconomic şi este în curs de reglementare în altele.

Cererea elaborării unor prevederi legale privind protecţia drepturilor profesionale a venit în urma constatării că metodele de investigare psihologică sunt practicate, într-un mod impropriu, de persoane lipsite de pregătire în domeniu, cu consecinţe negative asupra calităţii investigării. Activitatea de testare psihologică a fost cea mai afectată de imixtiunea neprofesioniştilor. În urma constatării că o serie de teste sunt elaborate şi folosite în mod superficial chiar de către unii psihologi slab pregătiţi sau rupţi de profesie (care nu au practicat-o timp îndelungat), s-au elaborat coduri deontologice, ca documente ale asociaţiilor profesionale ale psihologilor. Cele mai cunoscute sunt cele elaborate de APA, reevaluate periodic.

O parte dintre prevederile codurilor de deontologie sunt centrate pe problemele psihologiei aplicate şi pe cele legate de cercetare. Un exemplu simplu şi edificator de probleme intrate sub incidenţa codurilor de deontologie este următorul: anunţarea subiectului că a rezolvat greşit o problemă, într-un experiment sau într-o activitate de testare psihologică. Puţin câte puţin, s-a creat obişnuinţa de a furniza informaţii subiectului asupra a ceea ce s-a petrecut în timpul experienţei.

În Franţa, un comitet naţional de etică, alcătuit în mare parte din cercetători, se află la originea legilor care reglementează experimentarea asupra fiinţelor umane. Desigur, aceste legi au fost concepute să vizeze în principal experimentarea medicală, dar cuprind şi psihologia, care ridică probleme specifice.

Expunem câteva dintre principiile elaborate de comitetul francez:

- Subiecţii trebuie să fie voluntari.
- Ei nu trebuie să fie remuneraţi pentru activitatea lor.
- Înainte de examinare, subiecţilor li se cere clar consimţământul.

Interzicerea oricărei plăţi are drept scop evitarea constituirii de grupuri de „subiecţi profesionişti". Există reticenţe în a amesteca problema banilor în realizarea de cercetări fundamentale, considerate dezinteresate. Există, în schimb, şi o altă optică potrivit căreia se poate considera normală remunerarea celui care îşi consacră o parte din timpul său, mai ales că în acest fel se pot găsi uşor subiecţi. După susţinătorii acestei optici, lipsa motivaţiei financiare poate motiva unele comportamente artificiale şi ludice în rândul subiecţilor. Desigur, apare problema diferenţei de motivaţie între subiecţii plătiţi şi cei neplătiţi.

Cine poate aplica testele ? Este o întrebare firească, la care au răspuns anumiţi specialişti, dar al cărei răspuns nu-l cunosc cei care aplică testele psihologice fără a avea o pregătire de specialitate. Aşadar, care sunt persoanele şi ce calităţi trebuie să aibă acestea pentru a aplica şi exploata teste psihologice? Răspunsul este clar şi fără dubiu: specialistul destinat unor astfel de activităţi este psihologul, absolvent al unei forme de învăţământ universitar, atestat de o asociaţie profesională pentru a practica respectivele activităţi. Nu întotdeauna se întâmplă acest lucru, mai ales în România, ţară unde statutul psihologului nu este precizat legal. Ambiguitatea legislativă face ca, în acest domeniu, să se pronunţe sociologi, ingineri, filologi, economişti, medici etc. Astfel, rezultatele prezentate suferă de cele mai multe ori de simplism şi de lipsă de responsabilitate. La aceleaşi rezultate poate conduce prestaţia unui titrat universitar în psihologie care este rupt de profesie, adică a unei persoane care, după absolvire, nu a practicat psihologia o perioadă îndelungată de timp.

Nu putem clarifica valoarea şi limitele examenului psihologic prin teste, dacă nu facem referiri la statutul psihologului, la exigenţele formal juridice ale profesiei. O serie de ţări, mai ales europene, au făcut deja paşi absolut necesari în această privinţă. După informaţiile pe care le oferă Michael Hockel (1984), preşedinte, în 1982, a Federaţiei Europene a Asociaţiei Profesionale a Psihologilor (FEPPA), într-o conferinţă de la Edinborough (25 iulie 1982), prima ţară care a introdus în nomenclatoarele oficiale profesiunea de psiholog, pe baza atestării prin examen de stat, a fost Germania (în 1941). O urmează Franţa (în 1947). În 1948, se adoptă în SUA o primă lege a psihologului şi, în 1977, ultimele state americane introduc astfel de legi. Urmează Olanda, în 1971, Portugalia, în 1972, Norvegia, în 1973, Suedia, în 1978, Grecia şi Spania, în 1979 etc.

Michael Hockel formulează cinci puncte pentru descrierea identităţii profesionale a psihologului, puncte pe care le considerăm consistente:

1) psiholog este doar cel care a beneficiat de o formare în ştiinţa psihologiei. Hockel apreciază că în ultimii o sută de ani nu s-a făcut clar distincţia între psihologi şi nepsihologi (articolul în care face aceste aprecieri a apărut în anul 1984);
2) atitudinile de fundamentare cognitivă a ştiinţei psihologice corespund atitudinii ce trebuie pretinsă practicării profesionale a psihologiei;
3) psihologul este format ştiinţific pentru anticipare;
4) psihologul este obligat să respecte etica profesională;
5) practica profesional psihologică este în acelaşi timp şi practică de cercetare;

Dar aceste puncte ar putea fi continuate cu următoarele, stabilite de noi:

6) problematica psihologului trebuie să fie formulată ca problemă psihologică;
7) formularea ipotezelor trebuie să se facă în aşa fel încât să ducă la cunoaştere;
8) psihologul efectuează verificarea statistico-experimentală a ipotezelor;
9) psihologul colectează datele din practică şi procedează la evaluarea lor din perspectiva problemei vizate;
10) psihologul efectuează evaluarea finală şi, în cazul nereuşitei, trece la analizarea problemei.

Un aspect care merită evidenţiat este acela al preocupărilor asociaţiilor profesionale de a întocmi scale de apreciere pentru calitatea testelor psihologice. Astfel de grile există în mai multe modele, dar nu vom prezenta decât unul (*apud* Stoll, 1978, p. 322). Scala respectivă este prezentată în *NIP-Documentatie van tests en testresearch in Nederland*, document elaborat în 1969, de către Societatea Olandeză de Psihologie. Ea îşi propune să fie un ghid perfectibil pentru aprecierea valorii unui test şi cuprinde şase trepte descrescătoare în privinţa aprecierii, precizând mai multe criterii. Treapta A a scalei semnifică „foarte bine" şi treapta F „foarte rău". Totodată, aprecierea se face în funcţie de 12 criterii. Scala se referă la calităţile psihometrice ale testelor, deci la sensul lor restrâns. O astfel de scală oferă nota A unui test, dacă el îndeplineşte complet primele şase criterii de nivel foarte bun. Sunt prevăzute exigenţe pentru rangurile B, C, D, E şi F (ultimele două raguri acordate semnifică faptul că respectivele construcţii psihometrice nu mai fac parte din teste).

Tabelul 42

Scală de apreciere a calităţilor unui test

Scală de apreciere	Foarte bine	(...)				Foarte rău
	A	B	C	D	E	F
1. Sistem obiectiv de punctare şi sistem de decizie	X	X	X	X		
2. Norme pentru grupe definite	X	X	X			
3. Date pozitive despre fidelitate	X	X				
4. Nici o informaţie clară despre fidelitatea negativă	X	X				
5. Date suficiente despre validitatea de construct şi validitatea predictivă	X					
6. Puţine rezultate nevalide	X					
7. Date suficiente despre validitatea predictivă şi despre validitatea de construct; ambele sunt prezente, însă nu consistent		X				
8. Fidelitate incompletă, însă suficientă în cercetări ulterioare			X			
9. Câteva afirmaţii despre validitatea de construct şi predictivă; când nu există nici o indicaţie despre validitate – fidelitate pozitivă			X			
10. Norme pentru grupe definite sau rezultate pozitive despre fidelitate ori câteva rezultate pozitive despre validitate; mai multe din această ultimă categorie, care sunt însă inconsistente				X		
11. Nici o cercetare cunoscută					X	
12. Datele existente indică o construcţie deficitară a testului sau o teorie deficitară						X

Litera X în cadrul unei celule arată că în partea de sus a coloanelor se poate citi calitatea unui test care corespunde conţinutului afirmaţiilor criteriului din linie. Astfel la linia 12, X-ul semnifică faptul că testul este foarte rău. Acolo unde găsim completate mai multe celule ale unei linii cu literele X, se reliefează că testul care satisface conţinutul afirmaţiei din linie nu poate fi delimitat printr-un singur criteriu, ci prin mai multe. Astfel, la linia 1 se poate aprecia că un test care are un sistem obiectiv de punctare poate avea aprecieri de la A la D, satisfacerea altor criterii ajutând la o stabilire clară a calităţii testului. O astfel de scală trebuie completată de experţi. Anchete despre calitatea testelor, pe baza completării unor scale, sunt realizate periodic de o serie de asociaţii profesionale, cum ar fi Asociaţia profesională elveţiană pentru psihologie aplicată (Schweizerischer Berufsverband für Angewandte Psychologie - SBAP).

3. Teoria şi practica aplicării testelor în România

Deşi era o ştiinţă în constituire, este o evidenţă faptul că psihologiei i s-a acordat o atenţie specială încă de la apariţia primei universităţi româneşti. Disciplina *psihologie* figurează printre obiectele de studiu la prima Universitate din ţară, înfiinţată la Iaşi, în anul 1860, de către domnitorul Alexandru Ioan Cuza, şi la cea de-a doua Universitate, cea din Bucureşti. Profesorul Asaki, de la Universitatea din Bucureşti, se numără printre participanţii la primul Congres internaţional de psihologie. În domeniul construcţiei şi aplicării testelor psihologice, Nicolae Vaschide poate fi considerat un nume de referinţă, imediat după lansarea scalei de inteligenţă Binet-Simon. Preocupările de început în domeniul testelor psihologice au fost prezente în oraşele Iaşi, Bucureşti şi Cluj-Napoca, dar centrul de interes l-a reprezentat, fără îndoială, acesta din urmă, unde prin intermediul Institutului de psihologie s-au creat şi adaptat teste, iniţiativă preluată, până la izbucnirea celui de-al doilea război mondial, la dimensiuni mai mici, şi de Laboratorul de psihologie experimentală din Bucureşti.

În România, începutul unor preocupări sistematice în psihologia aplicată se face simţit în jurul anilor '30, mult mai târziu decât în ţările din vestul Europei, inclusiv încercări de etalonare şi adaptare a celor mai cunoscute teste de pe continent la acea vreme. În 1938, Florian Ştefănescu-Goangă înfiinţează *Revista de Psihologie*, care va publica ulterior contribuţii ale psihologilor români în diverse domenii, precum şi articole de comentarii asupra mişcării psihologice. Acelaşi autor, în lucrarea *Măsurarea inteligenţei*, apărută în 1940 la Editura Institutului Psihotehnic din Cluj-Napoca, revizuieşte, adaptează şi completează scala de inteligenţă Binet-Simon. I.M. Nestor traduce testele Buhler-Hetzer, având copiii mici ca destinaţie psihodiagnostică, şi etalonează testul Kohlmann-Anderson. Testul Pintner-Paterson (1972) este etalonat de către psihologi din Cluj-Napoca şi Iaşi. În perioada amintită, se înfiinţează oficii de orientare profesională şi se practică examene de selecţie într-o serie de întreprinderi industriale, în care se folosesc mai ales teste de aptitudini tehnice. În lansarea testelor de aptitudini, cea mai merituoasă personalitate este Nicolae Mărgineanu, autor care, în 1935, este publicat de *Journal of General Psychology* (nr. 12, pp. 397-415) cu un articol despre concepţia dinamică în psihologie (*apud* Lewin, 1975, p. 211). Nicolae Mărgineanu iniţiază în ţara noastră primele cercetări

psihologice în care era folosită metoda analizei factoriale. El realizează şi o baterie de teste despre gândirea tehnică, observarea unor forme şi figuri, gândirea spaţială. Concepţia sa despre examenul psihotehnic prin intermediul testelor era una flexibilă şi dinamică, susţinând în lucrarea *Psihotehnica* (Editura Institutului Psihotehnic, Sibiu, 1943) „aşezarea disciplinei psihotehnicii pe o nouă temelie, *psihologia persoanei*" (*apud* Nestor, 1974, p. 11). M. Peteanu realizează un grup de probe de aptitudini tehnice, iar D. Tudoran adaptează probe pentru diferite aspecte ale temperamentului. Un chestionar cu două forme paralele, atitudinea faţă de tradiţie şi atitudinea faţă de progres, realizează Anatoli Chircev, prevăzând trei forme de răspuns: afirmativ, negativ şi îndoielnic. Activităţi meritorii în planul elaborării instrumentelor de cercetare psihologică au Al. Roşca, S. Cupcea, L. Rusu.

Observăm, din trecerea în revistă a unor asemenea realizări, că şcoala românească de psihometrie a avut preocupări în pas cu evoluţia mişcării ştiinţifice europene dintre cele două războaie mondiale.

Tot în perioada interbelică, apar primele laboratoare psihotehnice. În anul 1926, se înfiinţează un laborator psihotehnic la Societatea de Tramvaie Bucureşti (STB), apreciat de Constantin Doru Blaj (1978, p. 17) ca fiind „destul de bine dotat pentru acea vreme". Testele creion-hârtie şi cele aparative au fost importate din Franţa. Conducătorii de tramvai sunt examinaţi după metodologia elaborată de J.M. Lahy pentru SNCF. Peste zece ani, Căile Ferate Române înfiinţează un laborator similar pentru nevoile proprii de examinare psihologică.

Perioada războiului şi a regimului totalitar, mai ales la începutul său, a reprezentat un hiatus, cu efecte inhibitorii asupra unei mişcări ştiinţifice desfăşurate într-o ambianţă de competenţă, seriozitate şi entuziasm intelectual. Nu înseamnă însă că asemenea preocupări au încetat cu totul.

În Iaşi, preocupări sistematice în domeniul psihometriei au avut psihologul cercetător Ioan Holban, care a studiat adaptarea testelor folosite în domeniul şcolar şi a activat în cadrul Institutului de ştiinţe pedagogice, precum şi profesorul universitar Andrei Cosmovici, din cadrul Catedrei de psihologie a Universităţii „Al.I. Cuza", preocupat de crearea unor probe psihologice noi, în care să îmbine analiza cantitativă cu cea calitativă, în ipoteza că determinarea calităţilor gândirii este mai importantă decât simpla înregistrare a rezultatelor cantitative. Crearea unei simptomatologii ştiinţifice, care să se refere la calităţile aptitudinilor, investigate prin intermediul testelor, a fost o constantă atitudinală în această problemă controversată. Profesorul Andrei Cosmovici este şi autorul unor teste psihologice utilizate în cercetarea ştiinţifică din domeniul psihologiei.

Dintre acestea, merită menţionat testul *Motive reale*, având un suport în teoria psihologului german Hans Thomae (1968) care, bazându-se pe analiza unui număr de peste 1.000 de autobiografii, a descris „temele existenţei", adică scopurile, motivele principale ale vieţii oamenilor. De fapt, este vorba chiar despre valorile care le ghidează existenţa. Andrei Cosmovici a creat un test proiectiv de interpretare vizând tocmai analizarea acestor motive.

Testul constă din 15 relatări variate privind comportamentul unor persoane. Apoi se indică între şase şi zece justificări prin care s-ar explica felul de a fi al personajului descris. Dintre ele subiectul trebuie să aleagă una, cea mai probabilă. De exemplu, itemul nr. VII din proba pentru băieţi (elevi sau studenţi), intitulată *Furtul*:

Traian avea un stilou foarte bun şi scump, cu care se mândrea. Zilele trecute, la ultima oră, căutând în geantă (unde îl ţinea de obicei), nu l-a mai găsit. Cu toate investigaţiile făcute de diriginte, hoţul n-a putut fi descoperit. Care dintre elevi a putut fi autorul acestui act?

- Poate că a fost un băiat cu multe greutăţi materiale, care s-a gândit că vânzându-l va putea mânca un timp alimente gustoase.
- Un coleg era invidios pe Traian, care, având părinţi cu venituri mari, are haine elegante şi multe lucruri scumpe.
- Un elev îşi dorea mult un stilou şi, de aceea, „l-a şterpelit".
- Un elev a furat ca să dovedească unor băieţi din cartier că este curajos şi îndemânatic.
- Cineva a vrut să se răzbune pe Traian şi a găsit acest mijloc.
- Un băiat a furat stiloul ca să-l dăruiască unei fete din cartier, căreia îi plac lucrurile deosebite.
- Alt motiv.

Argumentele au fost clasificate, în mod independent, de către trei psihologi, potrivit cu indicaţiile lui H. Thomae. Au fost reţinute zece explicaţii ale conduitei: 1) realizarea de sine; 2) afinitatea socială (sociabilitatea); 3) ridicarea socială (ambiţia); 4) normativă (obligaţie socială); 5) activarea existenţei (distracţiile); 6) materială; 7) erotismul; 8) anxietatea; 9) agresiunea; 10) compensarea.

Pentru fiecare explicaţie aleasă se atribuie un punct motivului corespunzător. Subiecţii obţin toţi 15 puncte distribuite foarte variat. Pentru repartizarea lor există o grilă de corectură. Un punct nu are aceeaşi valoare la toate motivele, fiindcă din totalul de 112 argumente prezentate găsim 20 de explicaţii încadrându-se în „realizarea de sine", 11 la „materială" şi doar 6 la „anxietate". Astfel, un punct valorează 0,05 la realizarea de sine, la materială 0,09, iar la anxietate 0,17. Deci, după stabilirea punctajului brut, cotele se ponderează şi devin comparabile între ele; în locul ponderării, se poate efectua o etalonare a fiecărui motiv luat separat. Existenţa cotelor numerice permite efectuarea de variate calcule statistice. Deseori, şi analiza calitativă este foarte edificatoare.

Pentru ca povestirile şi personajele să fie cât mai apropiate de lumea subiectului şi pentru a-i finaliza proiecţia, s-au elaborat patru variante: MRB pentru tineri studioşi, MRB pentru tinere, MRM pentru bărbaţi adulţi (în special muncitori) şi (avându-l ca autor pe Florian Vlăescu) varianta pentru femei adulte (în special muncitoare). Variantele diferă prin denumirile personajelor şi prin câteva relatări. Dar interpretarea urmează aceeaşi cale.

S-a verificat dacă cele zece motive sunt independente. Corelaţiile dintre ele s-au dovedit fie negative, fie nesemnificative statistic. Varianta elaborată de F. Vlăescu a evidenţiat o fidelitate (prin retestare) de 0,75, ceea ce este foarte bine pentru un test proiectiv.

Validitatea a fost confirmată de rezultatele concordante cu datele observaţiei, obţinute în şapte ample cercetări. Astfel, testând elevii cei mai indisciplinaţi dintr-un liceu, comparativ cu loturi de copii obişnuiţi, a reieşit că „indisciplina" corelează cu „agresivitatea" şi „erotismul", fiind prezentă şi o „insensibilitate la jigniri". Comparând cotele obţinute la vatmani la testele MRM cu numărul şi gravitatea accidentelor produse de-a

lungul anilor, a reieşit că ele sunt favorizate de „agresivitate", „activarea existenţei" (preocuparea pentru distracţii) şi „erotism". Iar bunul conducător de vehicule manifestă oarecare „anxietate" (în sensul întrevederii pericolelor) şi „ambiţie". Prezenţa fricii de accidente, este de fapt o componentă a prudenţei, factor esenţial în siguranţa circulaţiei şi foarte greu de diagnosticat.

Asemenea rezultate recomandă testele MRM într-o mare varietate de cazuri în care se cere un psihodiagnostic al însuşirilor de personalitate, fiindcă prezenţa unor factori ca agresivitatea, anxietatea, ambiţia şi erotismul se constată greu prin chestionare, unde intervin tendinţele de faţadă. Acest test proiectiv e mai puţin transparent, determinând indirect existenţa unor motive (valori) de prezenţa cărora subiectul poate fi conştient.

În adaptarea şi răspândirea unor teste de uz curent, în cadrul unor laboratoare de psihologie din Iaşi şi din întreaga ţară, un rol important l-a avut Serviciul de sănătate mintală din cadrul Spitalului de neuropsihiatrie din Iaşi, care, prin psihologul Nicolae Cosmovici, a „lansat" în folosinţă probe reputate pe plan mondial din domeniul investigării personalităţii şi aptitudinilor. Rolul profesorului Petre Brânzei (vezi 1974) în reconsiderarea activităţii psihologului clinician, şi nu numai a acestuia, este de necontestat.

Activitatea de examinare psihologică prin intermediul testelor psihologice a fost totuşi slab reprezentată, dacă ne comparăm cu ţările cu practică intensă în această privinţă, situaţie explicată prin lipsa testelor psihologice. Chiar la nivelul anului 2000, efectuarea unui test psihologic este o experienţă relativ rară pentru persoanele obişnuite. Prezenţa unui test sau a unei baterii de teste procurate de la producători avizaţi, cu situaţia criteriilor psihometrice bine pusă la punct, era relativ puţin frecventă. În majoritatea cazurilor, testele erau împrumutate de către psihologi de la colegi care activau în rarele instituţii bine dotate. Circulaţia probelor în acest mod se făcea cu o pierdere considerabilă de informaţie şi de adecvare.

În această ambianţă psihologică rarefiată în privinţa dotării cu teste psihologice, au existat totuşi laboratoare cu activitate de vârf şi psihologi foarte bine pregătiţi. Laboratoarele de excepţie în privinţa dotării şi competenţei au fost (unele există în continuare): Laboratorul de psihologie aeronautică, Laboratorul de psihologie al Regiei Autonome de Transport Bucureşti, Laboratorul central al MTTC, Laboratorul de psihologie al Combinatului chimic Făgăraş, Laboratoarele de psihologie ale Ministerului de Interne şi ale Ministerului Apărării Naţionale, Laboratorul de psihologie de la IPRS Băneasa, o mică parte dintre laboratoarele organizate pe lângă întreprinderile judeţene de transport în comun (Arad, Iaşi), Laboratorul de psihologie din cadrul Institutului de cercetări pedagogice, Laboratorul de psihologie al Combinatului Siderurgic din Reşiţa (care a beneficiat de prezenţa unui psiholog cu deosebite calităţi organizatorice, în persoana lui Zoltán Bogáthy)[1], o parte dintre laboratoarele de sănătate mintală, printre care şi cel de la Iaşi.

1. La Reşiţa a funcţionat primul Laborator psihotehnic uzinal din România, înfiinţat în anul 1938 de Nicolae Mărgineanu. Laboratorul a funcţionat până în anul 1948, când a fost desfiinţat din motive lesne de înţeles. Activitatea de psihologie a fost reluată în 1968. Zoltán Bogáthy spune în cartea sa *Îndreptarul psihologului industrial*, apărută la Editura Ştiinţifică şi Enciclopedică, în 1974: „Desigur, laboratorul uzinal de psihologie, înfiinţat în anul 1968, nu are nimic comun cu psihotehnica anului 1938" (p. 22). Precizarea ni se pare ca făcând parte dintr-o strategie de supravieţuire a activităţii psihologului şi psihologiei în România totalitară. Paginile 25-28 ale cărţii amintite cuprind liste ale unei dotări de excepţie cu aparatură şi teste creion-hârtie.

După cel de-al doilea război mondial, teoria psihometrică a avut reprezentanți de certă ținută științifică, dar care au activat în condiții de relativă izolare față de mișcarea ideilor și aplicațiilor psihologice pe plan mondial. În afară de autorii amintiți, Gheorghe Zapan s-a manifestat în câmpul cercetării științifice printr-o mare diversitate tematică, din care amintim statistica psihologică, psihologia experimentală, modelarea activității umane pe baza teoriei informației, diagnosticul sistemului temperamental, modernizarea aparaturii și tehnologiei de cercetare experimentală.

Profesorul Ioan Radu, de la Universitatea din Cluj-Napoca, are ca preocupări științifice constante acuratețea măsurării în psihologie, semnificația diferiților indicatori statistici și depășirea unor stereotipuri metodologice în cercetarea concretă. Referindu-se la o problematică frecvent abordată în metodologia cercetării psihologice, și anume cea a distribuției normale, după care se apreciază majoritatea performanțelor în psihometrie, Ioan Radu (1993 b) spune: „În practică se încearcă prin toate mijloacele *să se normalizeze* curbele. Iregularitățile histogramei trebuie să fie prea pronunțate pentru a fi luate în considerație. În felul acesta trec nesesizate fie suprapuneri de variabile, fie contopiri de grupuri eterogene în distribuții ușor bimodale".

Regretatul Valeriu Ceauşu este un alt reprezentant merituos al psihometriei românești. Calitățile sale științifice și organizatorice l-au recomandat pentru funcții de conducere în Asociația Psihologilor din România, unde a ocupat funcția de președinte, în perioada 1987-1990, și de vicepreședinte, în perioada 1981-1987. De asemenea, a ocupat funcțiile de șef al Laboratorului central CFR (1956-1957) și de șef al Laboratorului psihologic al aviației din cadrul Centrului de medicină astronautică, din 1959. În acest centru, activitatea de examinare psihologică, presupunând și utilizarea frecventă a testelor psihologice, s-a desfășurat la standarde mondiale. Unii dintre colaboratorii lui Valeriu Ceauşu, care au activat în acest centru, au devenit personalități ale psihologiei românești.

Pe lângă traducerea, construcția, adaptarea și validarea unor probe psihologice, Valeriu Ceauşu a avut preocupări de cercetare științifică vizând eliminarea surselor de eroare din măsurătorile psihologice, perfecționarea metodologiei investigației în scopul selecției profesionale și fundamentarea actului deciziei psihologice în urma unei examinări psihologice.

Mariana Roşca a analizat cu mult simț analitic evoluția structurală a testelor psihologice și a altor forme de investigare sistematică a caracteristicilor personalității, într-o lucrare de referință în literatura de specialitate românească, *Metode de psihodiagostic*, din 1972. În prefața lucrării (p. 5), autoarea spune: „Această lucrare mai urmărește formarea unei concepții în legătură cu responsabilitatea științifică și morală ce trebuie să stea la baza creării și utilizării metodelor standardizate de examinare psihologică, cu limitele fiecărui test, și ale diagnosticului psihologic în general cu principalele cauze ale erorilor".

La Universitatea din București, profesorul universitar Ursula Şchiopu a predat un curs de psihodiagnostic, preluat ulterior de profesorul universitar Nicolae Mitrofan, autor care, într-o lucrare recentă (*Testarea psihologică a copilului mic*, din 1997), face o analiză a psihodiagnosticului românesc până în 1991: „Concluziile care nu sunt deloc îmbucurătoare, am putea să le sistematizăm astfel: • lipsa aproape totală a preocupărilor românești privind teoria psihodiagnostică și psihometrică pe o perioadă de aproape 20 de ani; • absența unor lucrări românești publicate, cele mai vechi aparținând anilor '70 (de exemplu U. Şchiopu, *Introducere în psihodiagnostic*, TUB, 1976; M. Roşca,

Metode de cunoaștere a personalității, EDP, București, 1972; A. Cosmovici (coord.), *Metode pentru cunoașterea personalității*, EDP, București, 1971; A. Cosmovici, *Curs de psihologie diferențială și psihodiagnostic*, Univ. «Al.I. Cuza», Iași, 1974; T. Kulcsar, *Îndrumar psihodiagnostic*, vol. II, Univ. Cluj-Napoca, 1976.); • întârzieri nepermise din punct de vedere științific în revizuirea și reetalonarea unor instrumente psihodiagnostice folosite de către practicieni. În cel mai bun caz, pentru unele teste au fost folosite etaloane create în interiorul unor categorii socioprofesionale".

O realizare deosebită în adaptarea unei probe psihologice în România este cea a psihologului Cornel Cotor, fost șef al Laboratorului de psihologie din cadrul Centrului de organizare cibernetică în construcții. El a adaptat testul lui R.B. Cattell, *16 PF*, pentru mediul sociocultural românesc. De asemenea, a întocmit o bibliotecă de programe pentru prelucrarea statistică pe calculator a testelor psihologice[1].

După 1990, relansarea învățământului psihotehnic a creat bune condiții pentru revirimentul rapid al cercetării psihologice și, implicit, al dezvoltării unei activități de creare și dezvoltare a testelor psihologice.

1. Informații preluate din Ursula Şchiopu (coord.), *Dicționar de psihologie*, Editura Babel, București, 1997, p. 180.

În loc de concluzii

Reamintim faptul că această lucrare nu se.vrea o analiză exhaustivă a problematicii testului psihologic, ci doar o introducere în această problematică, abordând preponderent o serie de teme considerate de interes. Ne-am străduit să facem o prezentare la un nivel de accesibilitate extins, acolo unde s-a putut realiza acest lucru. Am dori ca, după lectura cărţii de faţă, să se ajungă la concluzia că testul psihologic este un instrument de investigaţie viabil, uşor remodelabil, modern, deoarece el ţine seama de comandamentele modernităţii, dintre care se detaşează clar eficienţa şi economicitatea. Controversele existente în istoria sa au generat noi concepţii despre fundamentele teoretice, despre menirea şi construcţia lui. Desigur, ne referim mai ales la aporturile profesioniste, deoarece în istoria testului au apărut şi relativ numeroase reacţii ostile venite din sfera politicului şi ideologicului. O serie de critici benefice la adresa sa sunt realizate, în bună parte, de neprofesionişti, de persoane prea puţin familiarizate cu aspectele tehnice ale construcţiei şi exploatării informaţionale a testului psihologic, care s-au dovedit adesea utile, deoarece practica aplicării acestui instrument duce la un feedback din zona interacţiunii sociale. Testul slujeşte şi la evaluarea persoanelor, activitate care conduce implicit la perfecţionarea culturală, profesională şi morală a fiinţei umane. Direcţiile de perfecţionare au ţinut cont permanent de rezultatele şi reacţiile domeniilor practice de aplicare.

Astăzi, domeniul testului psihologic se prezintă ca unul foarte diversificat, dinamic, pregătit să facă faţă provocărilor realităţii. Creatorii testului psihologic nu au rămas indiferenţi la progresele disciplinelor şi tehnologiilor informatice. Întrebarea care se pune frecvent este dacă maşinile inteligente pot aplica şi interpreta testele psihologice şi dacă pot elabora concluzii care să poată fi utile celui examinat. Părerea autorului acestei cărţi este că niciodată o maşină, oricâtă inteligenţă artificială ar cuprinde, nu poate fi generator autonom de decizii.

Evoluţia testului psihologic a scos în evidenţă şi calităţile celor ce l-au creat şi dezvoltat, personalităţi ştiinţifice care, pe lângă creativitate, s-au remarcat prin perseverenţă şi prin devoţiunea faţă de ideea perfecţionării umane. La progresul testului psihologic participă actualmente ansambluri de specialişti care pot comunica din ce în ce mai rapid, datorită facilităţilor oferite de tehnologiile informatice contemporane. Băncile de date contribuie din ce în ce mai intens la perfecţionarea testului prin informaţia stocată şi sistematizată pe care o oferă.

Forma de investigaţie promovată de testul psihologic a „contaminat" şi alte discipline care au aplicat-o cu succes în evaluarea cunoştinţelor. Răspândirea testărilor din varii domenii ne întăreşte convingerea că testul reprezintă un instrument de investigaţie specific civilizaţiei moderne şi satisface exigenţele actuale referitoare la cunoaşterea eficientă.

Nu trebuie să se uite niciodată faptul că testul furnizează, de multe ori, doar o parte din informaţia necesară unui act decizional şi că absolutizarea posibilităţilor sale constituie o gravă eroare. Conştientizarea limitelor testului este una dintre primele elemente deontologice care trebuie reţinute, alături de acela referitor la exploatarea sa superficială, ceea ce aduce grave deservicii psihologului şi practicării psihologiei, în general.

Legat de aspectul relevanţei informaţionale este şi aspectul calităţii profesionale a persoanelor care folosesc testul psihologic în diagnostic şi prognostic. În această privinţă, nu există nici un dubiu, psihologul titrat universitar este specialistul care poate asigura o folosire competentă şi responsabilă. Psihologia constituie o disciplină dinamică, dezvoltată într-un ritm deosebit de rapid, acumulările teoretice şi procedurile practice ale acesteia neputând fi însuşite decât prin specializare universitară şi postuniversitară. Cunoştinţele despre testul psihologic pot fi însuşite şi de alţi profesionişti care colaborează frecvent cu psihologul, dar care nu au competenţa să decidă pe baza informaţiilor furnizate de test.

Bibliografie

* * * (1992), *Grand dictionnaire de la psychologie*, Larousse, Paris.

* * * (1972), *Testul Pintner-Paterson*, uz intern, Institutul de Ştiinţe Pedagogice, Iaşi.

* * * (1986), *Berufsordnung für Psychologen*, Deutscher Psychologen Verlag, Bonn.

* * * (1992), *International Psychology. View from around the world*, University of Nebraska Press, Lincoln, Londra.

* * * (1997), *Les Éditions du Centre de Psychologie Appliquée*, ianuarie, Paris.

* * * (1962), *Stanford-Binet Intelligence Scale Terman Merill*, Houghton Mifflin Company, Boston.

* * * (1998), *Annuaire-Guide de la psychologie*, ediţia a IV-a (texte adunate şi prezentate de Philippe Duval şi Marie-Thérèse Estivill), Hommes et Perspectives/ Martin Média, Révigny.

Albu, Monica (1998), *Construirea şi utilizarea testelor psihologice*, Editura Clusium, Cluj-Napoca.

Albu, Monica; Pitariu, Horia (1993), *Proiectarea testelor de cunoştinţe şi examenul asistat de calculator*, Editura Casa Cărţii de Ştiinţă, Cluj-Napoca.

Allport, Gordon W. (1991), *Structura şi dezvoltarea personalităţii*, Editura Didactică şi Pedagogică, Bucureşti.

Amelang, Manfred (1999), „Zur Lage der Psychologie: Einzelaspekte von Ausbildung und Beruf unter besonderer Berücksichtigung der ökonomischen Implikationen psychologischen Handelns", in *Psychologische Rundschau* (1994), Hogrefe, 50. Jahrgang/Heft, pp. 2-13.

Anastasi, Anne (1968), *Psychological Testing*, Macmillan Publishing Co., Inc., New York.

Anzieu, Didier; Chabert, Catherine (1997), *Les méthodes projectives*, Presses Universitaires de France, Paris.

Arias, Rosario Martinez (1995), *Psicometria: Teoría de los tests psicologicos y educativos*, Editorial Sintesis S.A., Madrid.

Armstrong, Michael (1999), *A handbook of human resource management practice*, Kogan Page Limites, Londra.

Balint M.; Balint, E. (1976), *Technique psychothérapeutique en médecine*, Petite Bibliothèque Payot, Paris.

Bartram, David (1994), „Fidélité et validité", in Beech, John R.; Harding, Leonora, *Tests, mode d'emploi. Guide de psychométrie*, ECPA, Paris.

Beech, John R.; Harding, Leonora (1994), *Tests, mode d'emploi. Guide de psychométrie*, ECPA, Paris.

Benesch, Helmuth (1995), *Atlas de la psychologie*, Librairie Générale Française, Paris.

Bergson, Henri (1999), *Eseu despre datele imediate ale conştiinţei*, Editura Antet, Bucureşti.

Bernaud, J.L. (1998), *Les méthodes d'évaluation de la personnalité*, Dunod, Paris.

Bernier, J.J.; Pietrulewicz, B. (1997), *La psychométrie*, Gaëtan Morin Éditeur, Montréal, Paris, Casablanca.

Blaj, Constantin Doru (1978), *Psihologia conducătorului auto*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Bogáthy, Zoltán (1978), *Îndreptarul psihologului industrial*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Bonboir, Anna (1972), *La méthode des tests en pédagogie*, Presses Universitaires de France, Paris.

Bonnardel, R. (1995), *Les tests psychomoteurs et de réactions psychomotrices*, EAP, Éditions Scientifiques et Psychologiques, Issy-les-Moulineaux.

Bonnet, C. ; Kekenbosch, C. ; Hguyen-Xuan, C. ; Richard, A.F. ; Tijus, C.A. (1994), „Mesure et modélisation", in R. Ghilione şi J.F. Richard (coord.), *Cours de Psychologie. 4. Mesure et analyses*, Dunod, Paris.

Boon-Heckl, Ulrike (1983), „Das Selbstverstandnis des verkehrspsychologischen Gutachters unter Anvendung computergesteuerten Testens", in *Psychologie in Österreich*, noiembrie 1983, 5, Viena.

Boring, Edwin G. (1963), *History, Psychology & Science : Selected Papers*, John Wiley & Sons, Inc., New York, Londra.

Botez, C. (1971), „Utilizarea testelor pentru depistarea şi aprecierea aptitudinilor profesionale", in *Revista de Psihologie*, vol. 17, nr. 1, Editura Academiei, Bucureşti.

Brânzei, Petre (1975), *Itinerar psihiatric*, Editura Junimea, Iaşi.

Braunstein, Jean-François ; Pewzner, Évelyne (1999), *Histoire de la Psychologie*, Armand Colin/ HER, Paris.

Brickenkamp, Rolf (coord.) (1997), *Handbuch psychologischer und pädagogischer Tests*, Hogrefe Verlag für Psychologie, Göttingen, Berna, Toronto, Seattle.

Cadin, Loïc ; Guérin, Francis ; Pigeyre, Frédérique (1997), *Gestion des ressources humaines. Pratique et éléments de théorie*, Dunod, Paris.

Carroll, J.B. (1993), *Human Cognitive Abilities*, Cambridge University Press, Cambridge.

Carter, Philip (1998), *Teste de inteligenţă*, Aldo Press, Bucureşti.

Caruso, Mario (1987), „Probleme der Entwicklung von Softwareprodukten für die Psychodiagnostik", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.

Chamboulant, S. (1951), *Manuel de Psychologie Apliquée*, Payot, Paris.

Chelcea, Adina (1997), *Cunoaşterea de sine şi a celorlalţi*, Editura Ştiinţă şi Tehnică, Bucureşti.

Clocotici, Valentin ; Stan, Aurel (2000), *Statistică aplicată în psihologie*, Editura Polirom, Iaşi.

Cosmovici, Andrei (1980), „Metode de investigare a personalităţii", in Beniamin Zörgo (coord.), *Probleme fundamentale ale psihologiei*, Editura Academiei, Bucureşti.

Cosmovici, Andrei (1971), „Aplicarea testelor pentru determinarea nivelului de dezvoltare intelectuală", in *Revista de Psihologie*, vol. 17, nr. 1, Editura Academiei, Bucuresti.

Cosmovici, Andrei (1974), *Psihologie diferenţială teoretică*, Universitatea „Al.I. Cuza", uz intern.

Cosmovici, Andrei ; Havârneanu, Cornel ; Stan, Aurel (1992), „Diagnosticarea reacţiilor în situaţii conflictuale şi sub presiunea timpului", in *Revista de Psihologie*, nr. 2, Editura Academiei Române, Bucureşti.

Cronbach, L.J. (1970), *Essential of psychological testing*, Harper & Row, New York.

Cronbach, Lee J. (1979), „Validität", in Karl-Hermann Wewetzer (coord.), *Psychologische Diagnostik*, Wissenschaftliche Buchgesellschaft, Darmstadt.

Cronbach, L.J. ; Drenth, P.J. (coord.) (1972), *Mental test and cultural adaptation*, Mouton Publishers, Haga.

De Montmollin, Maurice (1972), *Les Psychopitres*, Presses Universitaires de France, Paris.

Delius, Lothar (1971), „Herz- und Kreislauf-verbundene psychophysische Syndrome", in *Entspannugs-therapie psychosomatischer Störungen*, CIBA, Basel.

Deri, Susan (2000), *Introducere în testul Szondi*, Editura Paideia, Bucureşti.

Dickes, Paul ; Tournois, Jocelyne ; Flieler, André ; Kop, Jean-Luc (1994), *La Psychométrie*, Presses Universitaires de France, Paris.

Dorofte, Ionel (1981), *Analiza şi predicţia performanţelor umane*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Duhamel, Claude (1992), *Le choix des tests de recrutement. Guide du professionel*, Les Éditions d'Organisation, Paris.

Enescu, Gheorghe (1985), *Dicţionar de logică*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Eysenck, Hans Jürgen (1998), *Teste de inteligenţă*, Editura Queen, Bucureşti.

Fancher, Raymond E. (1990), *Pioneers of Psychology*, W.W. Northon & Company, New York, Londra.

Farreny, Henry (1968), *Les systèmes expertes – principes et exemples*, Cepadeus Éditions, Toulouse.

Feger, Hans (1972), *Skalierte Informationsmenge und Eindrucksurteil*, Verlag Hans Huber, Berna, Stuttgart, Viena.

Flonta, Mircea (1976), „Natura şi dinamica ştiinţei în concepţia lui Th.S. Kuhn", prefaţă la lucrarea lui Thomas Kuhn, *Structura revoluţiilor ştiinţifice*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Fondarai, J.A. ; Micas M. ; Alma, I. (1993), „Utilisation d'une batterie de tests psychométriques informatisée pour la détection du comportement normal et pathologique des individus dans le domaine de l'appréhension de l'information", in *Les techniques psychologiques d'évaluation des personnes, Actes du Congrès International*, Paris, 25-27 mai, 1993, INETOP, EAP, EAP, Issy-les-Moulineaux.

Forner, Yan (1993), „Quelle évaluation de soi proposent les logiciels d'orientation?", in *Les techniques psychologiques d'évaluation des personnes, Actes du Congrès International*, Paris, 25-27 mai, 1993, INETOP, EAP, EAP, Issy-les-Moulineaux.

Fraisse, Paul (1967), „Évolution de la psychologie expérimentale", in Fraisse, Paul ; Piaget, Jean *Traité de Psychologie expérimentale*, PUF, Paris.

Franke, Gabriele Helga (1997), „Über die Moglichkeiten der computerunterstutzten Darbietung beim revidierten Freiburger Personlichkeitsinventar. Zwei experimentelle Studien", in *Zeitschrift für Experimentelle Psychologie*, 1997, XLIV, 2, pp. 332-356, Hogrefe Verlag, Göttingen.

Funke, Joachim (1987), „Experimentelle Analyse des Umgangs mit computersimulierten dynamischen System", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung : U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.

Gould, Stephen Jay (1984), *La mal-mesure de l'homme*, Éditions Ramsay, Paris.

Gredler, A. ; Arminger-Waldeck E. ; Mateja, I. ; Samitz-Ebner, U. (1984), „Das Tätigkeitsfeld des Psychologen in den beruflichen Rachabilitation", in *Psychologie in Österreich*, 4 Jahrgang, Juni, Viena.

Gregory, Robert J. (1996), *Psychological Testing*, Allyn and Bacon, Boston, Londra, Toronto, Sydney, Tokyo, Singapore.

Grubitzsch, Siegfried. ; Rexilius, Günther (coord.) (1979), *Psychologische Grundbegriffe. Mensch und Gesellschaft în der Psychologie. Ein Handbuch*, Rowohlts Enziklöpedie.

Gueguen, Nicolas (1997), *Manuel de statistique pour psychologues*, Dunod, Paris.

Guillevic, Ch. ; Vautier, S. (1998), *Diagnostic et tests psychologiques*, Nathan, Paris.

Guthke, J. ; Böttcher, H.R. ; Sprung, L. (coord.) (1991), *Psychodiagnostik*, Deutscher Verlag der Wissenschaften, Berlin.

Hagebock, Jürgen (1994), *Computerunterstutzten Testen*, Hogrefe Verlag für Psychologie, Göttingen.

Hansgen, Klaus Dieter (1987), „Vorstellung einer Psychodiagnostischen Testsystems auf Klein-computer Kg – 85/1 zur Leistungsdiagnostik (BVLD)", in *Neue Trends in der Psycho-diagnostik*, Herausgabekollektiv Leitung U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.

Hathaway, S.R. ; McKinley, J.C. (1963), *MMPI Saarbrucken*, Verlag Hans Huber, Berna, Stuttgart.

Hathersall, David (1983), *History of Psychology*, Random, Ohio State University.

Havârneanu, Cornel (2000), *Cunoaşterea psihologică a persoanei. Posibilităţi de utilizare a computerului în psihologia aplicată*, Editura Polirom, Iaşi.

Havârneanu, Cornel (2000), *Metodologia cercetării în ştiinţele sociale*, Editura Erota, Iaşi.

Havârneanu, Cornel ; Stan, Aurel (1993), „Perspectivele sistemelor EXPERT în cadrul tehnologiilor didactice", in *Buletinul Cabinetului Pedagogic*, nr. 5, Iaşi.

Hehlmann, Wilhelm (1968), *Wörterbuch der Psychologie*, Alfred Kroner Verlag, Stuttgart.
Heine, Thomas; Krause, Andreas; Nach, Bourdon (1987), „Wiersma (Paper Pencil-Bwt) und einer apparativen BWT – Variante (Combitest 2)", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.
Heiss, Robert (1979), „Die diagnostischen Verfahren in der Psychologie", in Karl-Hermann Wewetzer (coord.), *Psychologische Diagnostik*, Wissenschaftliche Buchgesellschaft, Darmstadt.
Helman, Zena (1995-1996), „Înnoirea Rorschach-ului şi aportul său în psihopatologie din perspectiva curentului fenomeno-structural", in *Analele Universităţii „Al.I. Cuza" din Iaşi*, vol. IV-V, Psihologie-Ştiinţele Educaţiei.
Hergenhahn, B.R. (1997), *An Introduction to the History of Psychology* (ediţia a III-a), Brooks/Cole Publishing Company, Pacific Grove, Albany.
Hilgard, Ernest R. (coord.) (1978), *American Psychology in Historical Perspective*, American Psychological Association Inc., Washington.
Hilgard, R. Ernest (1987), *Psychology in America. A Historical Survey*, Harcourt Brace Jovanovich, Publishers, San Diego, New York, Chicago, Austin.
Hockel, Michael (1984), „Was ist und was kann ein Psychologe", in *Psychologie in Österreich*, nr. 3, Viena.
Hofstätter, Peter R. (1971), *Psychologie*, Fischer Taschenbuch Verlag, Frankfurt pe Main.
Holban, Ion (1995), *Testele de cunoştinţe*, Editura Didactică şi Pedagogică, R.A., Bucureşti.
Holban, Ion (1974), „Orientarea şcolară şi orientarea profesională, forme ale acţiunii de asistenţă complexă a persoanei", in Petre Brânzei (sub redacţia), *Adolescenţă şi adaptare*, Centrul de cercetări pentru problemele tineretului, Iaşi.
Horst, Paul (1980), „Eine Übersicht über die Gründzüge der Methoden der multivariaten Analyse", in Cattell, Raymond (coord.) *Handbuch der multivariaten experimentellen Psychologie*, Fachbuchhandlung für Psychologie, Verlagsabteilung, Frankfurt pe Main.
Horst, Paul (1971), *Messung und Vorhersahe. Eine Einfuhrung in die psychologische Testtheorie*, Verlag Julius Beltz, Weinheim, Berlin, Basel.
Hothersal, David (1983), *History of Psychology*, Random, Ohio State University.
Hoyos, Carl (1968), *Psychologie de la circulation routière*, Presses Universitaires de France, Paris.
Huber, Wolfram (1987), „Einsatz von Kleincomputern in der Entwicklungspsychologischen Arbeit", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.
Huteau, M.; Lautrey, J. (1995), *Évaluer l'intelligence*, Presses Universitaires de France, Paris.
Jacob, François (1972), *Logica viului*, Editura Enciclopedică Română, Bucureşti.
Jiménez, Paulino; Raab, Erich (1999), „Das Berufsbild von Psychologen im Vewrgleich mit anderen Berufen", in *Psychologische Rundschau*, Hogrefe, 50. Jahrgang/Heft 2, pp. 26-32.
Jungermann H.; Schultz, H. (1995), „Interaktive Programe zur Diagnose des Entscheidungsverhaltens", in Werner Sarges (coord.), *Management Diagnostic*, Hogrefe Verlag für Psychologie, Göttingen, Berna, Toronto, Seattle.
Katzbeck, Cristian; Kastner-Koller, Ursula; Deimann, Pia (1998), „Konstruktion und erste Schritte zur Konstruktvalidierung eines tretement-orientierten Kurzzeit-Lerntests für das Vorschulalter", in *Diagnostica*, 44 Heft 2, 100-111, Hogrefe Verlag, Göttingen.
Kline, Paul (1995), „La construction des tests", in Beech, John R.; Harding, Leonora, *Tests, mode d'emploi. Guide de psychométrie*, ECPA, Paris.
Klink, Dorothea (1998), „Papier-Bleistift-versus computerunterstutzte Administration kognitiver Fahigkeitstest: Eine Studie zur Aquivalenzfrage", in *Diagnostica Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie*, Hogreffe Verlag, 44 Jahrgang/Heft 2.
Kluwe, R.H. (1995), „Computergestutzte Systemsimulationen", in Werner Sarges (coord.), *Management Diagnostic*, Hogrefe Verlag für Psychologie, Göttingen, Berna, Toronto, Seattle.

Kristof, Walther (1983), *Klassische Testtheorie und Testkonstruktion*, in Feger, Hubert; Bredenkamp, Jürgen, *Messen und Testen*, Verlag für Psychologie, Dr. C.J. Hogrefe, Göttingen, Toronto, Zürich.
Kuhn, Thomas S. (1976), *Structura revoluţiilor ştiinţifice*, Editura Ştiinţifică şi Enciclopedică Bucureşti.
Lang, Alfred (1978), „Diagnostik und Autonomie der Person", in Urs Pulver; Alfred Lang; Fred W. Schmid (coord.), *Ist Psychodiagnostik verantwortbar ?*, Verlag Hans Huber, Berna, Stuttgart, Viena.
Larivée, Serge; Normandeau, Sylvie; Parent, Sophie (1996), „La filière francophone de la psychologie développementale différentielle", in *L'Année psychologique*.
Laveault, Dany; Grégoire, Jacques (1997), *Introduction aux théories des tests en sciences humaines*, De Boeck & Larcier, De Boeck Université, Paris, Bruxelles.
Leahey, Thomas H. (1980), *A History of Psychology*, Prentice Hall, Inc., Englewood Cliffs, New Jersey.
Legewie, Heiner; Ehlers, Wolfram (1978), *Knaurs Moderne Psychologie*, Droemer Knaur, München, Zürich.
Legrand, Michel (1979), *Léopold Szondi, son test, sa doctrine*, Pierre Mardaga Éditeur, Bruxelles.
Lévy-Leboyer, Claude (1995), *Évaluations et carrieres*, Dunod, Paris.
Lévy-Leboyer, Claude (1974), *Psychologie des organisations*, Dunod, Paris.
Lewin, Kurt (1975), *Psychologie dynamique. Les relations humaines*, PUF, Paris.
Lienert, Gustav A. (1967), *Testaufbau und Testanalyse*, Verlag Julius Beltz, Weinheim/Berlin.
Lukesh, Helmuth (1968), „Psychologischen Testverfahren", in Angela Schorr (coord.), *Handbuch der angewandte Psychologie*, Deutschen Psychologen Verfahren, Bonn.
Lussato, Ariane (1998), *Les tests de recrutement*, Presses Universitaires de France, Paris.
Macaire, Jean Pierre (1994), „Les techniques psychologiques d'évaluation des personnes à la SNCF", in *Techniques psychologiques d'évaluation. Actes du Congrès International*, Paris, 25-27 mai 1993, INETOP, EAP, EAP, Paris.
Magnusson, David (1975), *Testtheorie*, Verlag Frank Deuticke, Viena.
Mărgineanu, Nicolae (1973), *Condiţia umană*, Editura Ştiinţifică, Bucureşti,
Mattarazzo, Josef (1986), „Computerized Clinical Psychological Test Interpretation", in *American Psychologist*, vol. 41, ianuarie.
McCullogh, Leigh; Farrel, Albert; Langabough, Richard (1986), „The Development of a Microcomputer-Based Mental Health Information System", in *American Psychologist*, nr. 2, februarie.
Michel, Lothar; Conrad, Wolfgang (1982), „Theoretische Grundlagen psychometrischer Tests", in Dr. Karl-Josef Groffmann şi Dr. Lothar Michel, *Grundlagen psychologischer Diagnostik*, Enzyklopedie der Psychologie, Band I, Verlag für Psychologie, Dr. C.J. Hogrefe, Mannheim, Göttingen, Toronto, Zürich.
Miclea, M. (1991), „Psihologia cognitivă şi inteligenţa artificială", in I. Radu (coord.), *Introducere în psihologia contemporană*, Editura Sincron, Cluj-Napoca.
Miclea, Mircea (1994), *Psihologie cognitivă*, Casa de Editură Gloria S.R.L., Cluj-Napoca.
Mîrza, Alina; Matei, Liviu (1989), „Examinarea psihologică asistată de calculator cu chestionarul de personalitate Eysenck (E.P.Q.)", in *Revista de psihologie*, vol. 35, nr. 4, pp. 345-349, Bucureşti, octombrie-decembrie.
Mitrofan, Nicolae (2001), „Psihometria şi direcţiile ei de dezvoltare la început de mileniu", in Mielu Zlate (coord.), *Psihologia la răspântia mileniilor*, Editura Polirom, Iaşi.
Mitrofan, Nicolae (1997), *Testarea psihologică a copilului mic*, Edit Press Mihaela, Bucureşti.
Moser, Klaus (1999), „Selbsturteilung beruflicher Leistung: Überblick und offene Fragen", in *Psychologische Rundschau*, Hogrefe, 50. Jahrgang/Heft 1, pp. 14-25.
Mucchielli, A. (coord.) (1996), *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*, Armand Colin/Masson, Paris.

Mucchielli, Roger (1992), *L'examen psychotechnique*, ESF Éditeur, Paris.
Mueller, Fernand-Lucien (1968), *Histoire de la Psychologie de l'Antiquité à nos jours*, Payot, Paris.
Neculau, Adrian (1996), *29 de teste pentru a te cunoaşte*, Editura Polirom, Iaşi.
Nestor, I.M. (1974), *Psihologie industrială. Consideraţii practice de organizare*, Editura Politică, Bucureşti.
Nicolas, Serge (1999), „Les origines de la psychologie cognitive", in *Psychologie cognitive*, Bréal, col. „Grand Amphi", Paris.
Oriol, François ; Alferez, Emma (1990), „Le test a-t-il des rides", in *L'Enjeu humain de l'entreprise*, Créations Éditions Productions Publicitaires, Paris.
Parot, Françoise ; Richelle, Marc (1995), *Introducere în psihologie*, Editura Humanitas, Bucureşti.
Pavelcu, Vasile (1972), *Drama Psihologiei*, Editura Didactică şi Pedagogică, Bucureşti.
Pekrun, Reinhard (1993), „Geschichte von Differentieler Psychologie und Personlichkeitspsychologie", in Dr. Kurt Pawlik (coord.), *Grundlagen und Methoden der Differentiellen Psychologie*, Enzyklopedie der Psychologie, Themenbereich C, Serie VIII, Band 1, Hogrefe, Verlag für Psychologie, Hamburg, Göttingen, Berna, Toronto, Seattle.
Pichot, Pierre (1994), *Les tests mentaux*, Presses Universitaires de France, Paris.
Piéron, Henry (coord.) (1955), *Vocabulaire de la Psychologie*, Presses Universitaires de France, Paris.
Pitariu, H. ; Albu, M. (1989), „O aplicaţie informatică pentru testări psihologice : PSIGEN", in *Revista de psihologie*, vol. 37, nr. 3-4, pp. 205-208, Bucureşti, iulie-decembrie.
Pitariu, H.D. ; Albu, M. (1992), „Inventarul psihologic California : prezentare şi rezultate experimentale", in *Revista de psihologie*, vol. 39, nr. 3, pp. 235-244, Bucureşti, iulie-septembrie.
Pitariu, Horia (1983), *Psihologia selecţiei şi formării profesionale*, Editura Dacia, Cluj-Napoca.
Popper, Karl R. (1981), *Logica cercetării*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.
Postel, Jacques (1993), *Dictionnaire de Psychiatrie et de psychopathologie clinique*, Larousse, Paris.
Pulver, Urs (1978), „Autonomie, Wilkür und Objectivität", in Urs Pulver, Alfred Lang, Fred W. Schmid (coord.), *Ist Psychodiagnostik verantwortbar ?*, Verlag Hans Huber, Berna, Stuttgart, Viena.
Pulver, Urs (1978), „Die Krise der psychologischen Diagnostik – eine Koartationkrise", in Urs Pulver ; Alfred Lang ; Fred W. Schmid (coord.), *Ist Psychodiagnostik verantwortbar ?*, Verlag Hans Huber, Berna, Stuttgart, Viena.
Rader, Elisabeth (1987), „Kontenvalide Itemkonstruktion fur einen computergestutzten adaptativen Lerntest", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin, pp. 77-82.
Radu, I. (1991), „Metode de cercetare în psihologie", in I. Radu (coord.), *Introducere în psihologia contemporană*, Editura Sincron, Cluj-Napoca.
Radu, I. (1971), „Metoda testelor – văzută din perspectiva statisticii psihologice", in *Revista de Psihologie*, vol. 17, nr. 1, Editura Academiei, Bucureşti.
Radu, I. (1993), „Principii metodologice în elaborarea şi utilizarea probelor psihologice", in Radu, Ioan ; Miclea, Mircea ; Albu, Monica ; Moldovan, Olga ; Nemeş, Sofia ; Szamosközi, Stefan, *Metodologie psihologică şi analiza datelor*, Editura Sincron, Cluj-Napoca.
Radu, I. (1993b), „Depăşirea unor stereotipuri metodologice în cercetarea concretă", in Radu, Ioan ; Miclea, Mircea ; Albu, Monica ; Moldovan, Olga ; Nemeş, Sofia ; Szamosközi, Stefan, *Metodologie psihologică şi analiza datelor*, Editura Sincron, Cluj-Napoca.
Reber, Arthur S. (1985), *Dictionary of Psychology*, Penguin Books, Londra.
Remplein, Heinz (1975), *Psychologie der Persönlichkeit. Ein Lehrbuch der Persönlichkeitspsychologie*, Ernst Reinhardt Verlag, München/Basel.
Rettig, K. ; Hornke L.F. (1995), „Adaptativen Testen", in Werner Sarges (coord.), *Management Diagnostic*, Hogrefe Verlag für Psychologie, Göttingen, Berna, Toronto, Seattle.
Reuchlin, Maurice (1999), *Psihologie generală*, Editura Ştiinţifică, Bucureşti.
Reuchlin, Maurice, *Évolution de la psychologie différentielle*, Presses Universitaires de France, Paris, 1999.

Reuchlin, Maurice (1967), *Histoire de la Psychologie*, Presses Universitaires de France, Paris.
Reuchlin, Maurice (1992), *Introduction à la recherche en psychologie*, Éditions Nathan, Paris.
Reuchlin, Maurice (1962), *Les méthodes quantitatives en psychologie*, Presses Universitaires de France, Paris.
Reuchlin, Maurice (1976), *Précis de statistique*, Presses Universitaires de France, Paris.
Richard, J.F. ; Ricordeau, L. (1994), „La méthode des tests", in R. Ghiglione şi J.F. Richard (coord.), *Cours de Psychologie*, Dunod, Paris.
Richard, Meili (1978), „Bemerkungen zur sogenannten Krise der psychologischen Diagnostik", in Pulver, Urs ; Lang, Alfred ; Schmid, Fred W. (coord.), *Ist Psychodiagnostik verantwortbar ?*, Verlag Hans Huber, Berna, Stuttgart, Viena.
Robinson, Daniel H. (1970), *An Intellectual History of Psychology*, MacMillan Publishing Co., Inc., Collin MacMillan Publishers, Londra.
Robock, Abraham A. (1980), *Welgeschichte des Psychologie und Psychiatrie*, Walter-Verlag, Olten und Freiburg im Breisgau.
Roch, Jean (1993), „Équivalence des tests papier-crayon et des tests informatisés de capacités cognitives", in *Les techniques psychologiques d'évaluation des personnes, Actes du Congrès International*, Paris, 25-27 mai 1993, INETOP, EAP, EAP, Issy-les-Moulineaux, pp. 215-219.
Rorschach, Hermann (2000), *Manual de psihodiagnostic*, Editura Trei, Iaşi.
Roşca, Mariana (1972), *Metode de psihodiagnostic*, Editura Didactică şi Pedagogică, Bucureşti.
Roşca, Mariana (1971), „Valoarea şi limitele metodelor de diagnostic al personalităţii", in *Revista de Psihologie*, vol. 17, nr. 1, Editura Academiei, Bucureşti.
Rost, John ; Golombock, Susan (1989), *Modern Psychometrics. The Science of Psychological Assessment*, Routledge, Londra, New York.
Rost, Jürgen (1999), „Was ist aus dem Rasch-Model geworden ? ", in *Psychologische Rundschau*, Hogrefe, 50. Jahrgang/Heft 3.
Rotariu, Traian (coord.) ; Bădescu, Gabriel ; Culic, Irina ; Mezei, Elemér ; Mureşan, Cornelia (1999), *Metode statistice aplicate în ştiinţele sociale*, Editura Polirom, Iaşi.
Roulin, Lean-Luc (1990), „Informatisation des épreuves psychotechniques. Psychométrie et informatique", in *Informatique et Psychologie*, Presses Universitaires de Lyon, Lyon.
Rude, Nathalie ; Retel, Olivier (2000), *Statistique en psychologie*, In Press Éditions, Clamecy.
Saharian, William S. (1968), *History and Systems of Psychology*, FE Peacock Publisher, Inc., Itasca, Illinois.
Schlegel, Justin (1994), „L'évaluation dans les codes de déontologie", in *Les techniques psychologiques d'évaluation des personnes à la SNCF, Actes du Congrès International*, Paris, 25-27 mai 1993, INETOP, EAP, EAP, Issy-les-Moulineaux.
Schonman, Wooter (1989), „Testarea psihologică la Căile Ferate Olandeze", in *Revista de Psihologie*, nr. 3, vol. 35, Editura Academiei, Bucureşti.
Schorr, Angela (coord.) (1993), *Handwörterbuch der Angewandten Psychologie : die Angewandte Psychologie in Schlüsselbegriffen*, Deutscher Psychologen Verlag, Bonn.
Schufried, Garnot (1987), „Neue Tendenz in der Entwicklung psychotechnischer Geräte", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin, pp. 195-199.
Siewert, Horst H. (2000), *Teste de aptitudini profesionale şi rezolvarea lor*, Editura Gemma Pres, Bucureşti.
Siewert, Horst H. (1998), *Cum să ne calculăm coeficientul de inteligenţă*, Editura Gemma Press, Bucureşti.
Siewert, Horst H. ; Siewert, Renate (2000), *Testele pot fi trecute. Program de pregătire intensive*, Editura Gemma Press, Bucureşti.
Sporli, Siro (1978), *Kritische Theorie diagnostischer Praxis-dargestellt am Beispiel Verkehrspsychologie*, Verlag Hans Huber, Berna, Stuttgart, Viena.
Spreen, O. (coord.) (1963), *MMPI Saarbrucken*, Verlag Hans Huber, Berna, Stuttgart.

Sprung, Lothar; Sprung, Helga (1987), *Grundlagen der Methodologie und Methodik der Psychologie*, VEB Deutscher Verlag der Wissenschaften, Berlin.

Stan, Aurel (1999), „De la presistematic la sistematic în construcţia testului psihologic", in *Analele ştiinţifice ale Universităţii „Al.I. Cuza" din Iaşi* (serie nouă), vol. VIII, Editura Universităţii „Al.I. Cuza", Iaşi.

Stan, Aurel (2000), „Normele şi normarea în psihologie", in *Psihologia socială*, nr. 5, Universitatea „Al.I. Cuza", Editura Polirom, Iaşi.

Stan, Aurel; Havârneanu, Cornel (1989), „Utilizarea calculatorului în examinarea psihologică", in *Revista de Psihologie*, Editura Academiei, Bucureşti.

Stoll, François (1978), „Testgebrauch in die schweizerische Beratungpraxis", in Pulver, Urs; Lang, Alfred; Schmid, Fred (coord.), *Ist Psychodiagnostik verantwortbar ?*, Verlag Hans Huber, Berna, Stuttgart,Viena.

Stoloff, M.L.; Couch, J.V. (1987), *Computer use in psychology. A directory of Software*, American Psychological Association, Washington, DC.

Szamosközi, Stefan (1997), „Psihometria clasică şi evaluarea formativă", in *Cogniţie, creier, comportament*, vol. I, nr. 2, iunie, Cluj-Napoca.

Şchiopu, Ursula (coord.) (1997), *Dicţionar de psihologie*, Editura Babel, Bucureşti.

Thomae, Hans (1968), *Das Individum und seine Welt*, Hogrefe, Göttingen.

Tiffin, Joseph; McCormick, Ernest J. (1967), *Psychologie industrielle*, Presses Universitaires de France, Paris.

Trebici, Vladimir (coord.) (1985), *Mică enciclopedie de statistică*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Van der Veer, Gerrit (1990), „Aspecte vizuale ale interfeţei cu calculatorul – o privire psihologică", in *Revista de psihologie*, 3-4, vol. 36, Editura Academiei Române, Bucureşti.

Vasilescu, I.P. (1991), „Niveluri de complexitate ale asistenţei acordate de calculator în investigaţia psihodiagnostică", in *Revista de psihologie*, vol. 37, nr. 3-4, pp. 129-134, Bucureşti, iulie-decembrie.

Vukovich, A, F. (1964), „Die Konstruktion psychologischer Testen", in R. Heiss (coord.), *Psychologische Diagnostik*, Verlag für Psychologie Dr. C Hogrefe, Göttingen.

Weber, Peter; Leo, Harald; Regel, Hans (1987), „Combi 2 – ein Geratesyatem für die computergestutzte Psychodiagnostik", in *Neue Trends in der Psychodiagnostik*, Herausgabekollektiv Leitung, U. Schaarschmidt, Psychodiagnostisches Zentrum HUB, Berlin.

Weil-Barais, A. (coord.) (1997), *Les Méthodes en Psychologie*, Bréal, Rosny.

Wertheimer, Michael (1971), *Kurze Geschichte der Psychologie*, R. Piper & Co Verlag, München.

Zapan, Gheorghe (1984), „Teoria orientării şi selecţiei profesionale", in *Cunoaşterea şi aprecierea obiectivă a personalităţii*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Zazzo, René; Galifret-Granjon, N.; Hurtig, M.C.; Mathon, T.; Pecheux, M.G.; Santucci, H.; Stambak, M. (1977-1992), *Manuel pour l'examen psychologique de l'enfant*, Delachaux et Niestlé S.A., David Perret Éditeur, Neuchâtel, Paris.

Zăpârţan, Marioara (1990), *Eficienţa cunoaşterii factorilor de personalitate în orientarea şcolară şi profesională a elevilor*, Editura Dacia, Cluj-Napoca.

Zörgo, B. (1971), „Criteriile unui test ştiinţific", in *Revista de Psihologie*, vol. 17, nr. 1, Editura Academiei, Bucureşti.

Collegium. Psihologie

au apărut:

Adrian Neculau (coord.) – *Psihologie socială. Aspecte contemporane*

Andrei Cosmovici – *Psihologie generală*

W. Doise, J.-C. Deschamps, G. Mugny – *Psihologie socială experimentală*

Adrian Neculau, Gilles Ferréol (coord.) – *Minoritari, marginali, excluşi*

Constantin Cucoş – *Minciună, contrafacere, simulare. O abordare psihopedagogică*

Mielu Zlate (coord.) – *Psihologia vieţii cotidiene*

R.Y. Bourhis, J.-P. Leyens (coord.) – *Stereotipuri, discriminare şi relaţii intergrupuri*

Serge Moscovici – *Psihologia socială sau maşina de fabricat zei*

Adrian Neculau (coord.) – *Câmpul universitar şi actorii săi*

Adrian Neculau (coord.) – *Psihologia câmpului social. Reprezentările sociale*

J. Barus-Michel, F. Giust-Desprairies, Luc Ridel – *Crize. Abordare psihosocială clinică*

Andrei Cosmovici, Luminiţa Iacob (coord.) – *Psihologie şcolară*

Şerban Ionescu – *Paisprezece abordări în psihopatologie*

Serge Moscovici (coord.) – *Psihologia socială a relaţiilor cu celălalt*

Willem Doise, Gabriel Mugny – *Psihologie socială şi dezvoltare cognitivă*

Ana Stoica-Constantin, Adrian Neculau (coord.) – *Psihosociologia rezolvării conflictului*

André Sirota – *Conduite perverse în grup*

Adrian Neculau, G. Ferréol (coord.) – *Psihosociologia schimbării*

Mircea Miclea – *Psihologie cognitivă. Modele teoretico-experimentale*

Mielu Zlate – *Psihologia mecanismelor cognitive*

Gilles Ferréol (coord.) – *Identitatea, cetăţenia şi legăturile sociale*

Mielu Zlate – *Introducere în psihologie*

Ion Dafinoiu – *Elemente de psihoterapie integrativă*

Eva Drozda-Senkowska – *Psihologia socială experimentală*

Cornel Havârneanu – *Cunoaşterea psihologică a persoanei. Posibilităţi de utilizare a computerului în psihologia aplicată*

Steve Duck – *Psihologia relaţiilor interumane. Teorie, concepte şi noţiuni fundamentale*

Valentin Clocotici, Aurel Stan – *Statistică aplicată în psihologie*

Gilles Ferréol (coord.) – *Adolescenţii şi toxicomania*

Adrian Neculau (coord.) – *Analiza şi intervenţia în grupuri şi organizaţii*

Pierre De Visscher, Adrian Neculau (coord.) – *Dinamica grupurilor. Texte de bază*

Mihaela Roco – *Creativitate şi inteligenţă emoţională*

Jean Delacour – *Introducere în neuroştiinţele cognitive*
Ion Dafinoiu – *Personalitatea. Metode de abordare clinică : observaţia şi interviul*
Septimiu Chelcea – *Un secol de cercetări în psihosociologie*
Aurel Stan – *Testul psihologic. Evoluţie, construcţie, aplicaţii*

în pregătire :

Dan Goglează – *Întâlnire cu un psihoterapeut în practica schimbării. Exemple, cazuri, soluţii, comentarii*
Alex Mucchielli – *Arta de a influenţa*
Jean Claude Abric – *Psihologia comunicării*
A. Neculau, G. Ferréol (coord.) – *Violenţa. Aspecte psihosociale*

www.polirom.ro

Redactor : Adina Cobuz
Coperta : Manuela Oboroceanu
Tehnoredactor : Constantin Mihăescu

Bun de tipar : februarie 2002. Apărut : 2002
Editura Polirom, B-dul Copou nr. 4 • P.O. Box 266, 6600, Iaşi
Tel. & Fax (032) 21.41.00 ; (032) 21.41.11 ;
(032) 21.74.40 (difuzare) ; E-mail : office@polirom.ro
B-dul I.C. Brătianu nr. 6, et. 7 • P.O. BOX 1-728, 70700, Bucureşti
Tel. : (01) 313.89.78, E-mail : polirom@dnt.ro